

The Effect of Area Level Deprivation on Obesity in New Zealand: Analysis of The New Zealand Health Surveys

A thesis submitted in partial fulfilment of the requirements
for the Degree of Master of Health Sciences
in the University of Canterbury

by Andrew Adiguna Halim (ID: 62376311)

University of Canterbury

July 30, 2017

Contents

Acknowledgement	ix
Abstract	x
Glossary	xi
1 Background	1
2 Literature Review	4
2.1 Health Risks Associated with Obesity	4
2.2 Defining Overweight and Obesity	5
2.3 Lifestyle Interventions	7
2.4 The Genetics of Obesity	8
2.5 Agriculture	9
2.6 Urbanisation	10
2.7 Food Environment	11
2.8 Food Price and Consumption	12
2.9 Sugar-Sweetened Beverage and Fast Food Consumption on Body Weight	13
2.10 Sedentary Environment	13
2.11 Socio-Economic Status	14
2.12 Smoking and Obesity	15
2.13 Healthy Immigrant Effect	16
2.14 Literature Review Summary	16
3 Population and Methods	18
3.1 Theoretical Framework	18
3.2 Data	19
3.3 Hypotheses	20
3.4 Variables	21
3.4.1 Outcome Variables	21
3.4.2 Explanatory Variables	21

3.4.3	Covariates	21
3.5	Data Analysis	23
3.5.1	Data Cleaning	23
3.5.2	Descriptive and Bivariate Analyses	24
3.5.3	Regression Modelling	25
4	Results	26
4.1	Descriptive Analysis of The NZHS	26
4.2	Bivariate Analyses	30
4.3	Obesity Prevalence across Ethnic Groups	33
4.4	Regressions Result	34
4.5	Proportional Odds Assumption	38
5	Discussion	39
5.1	Summary of the findings	39
5.2	Area-Level Deprivation	39
5.3	Household Income Level	40
5.4	Obesity Trend	40
5.5	Ethnic Differences and The Universal BMI Cut-offs	41
5.6	Diet and Physical Activity	42
5.6.1	Fruit and Vegetable Guidelines	42
5.6.2	Fast Food and Soft Drink Consumption in Children	43
5.6.3	Physical Activity	43
5.7	Smoking Status	44
5.8	Healthy Immigrant	44
5.9	Urbanisation	45
5.10	Study Strengths and Limitations	45
6	Conclusions	47
6.1	Implications of Socio-Economic Status and BMI outcomes	47
6.2	Implications of The Universal BMI Cut-Offs	47
6.3	Implications of Diet and Physical Activity Interventions	48
6.4	Recommendations	48
	Appendix A R Codes	64

Appendix B Bivariate Tables	232
B.1 Bivariate Tables for The Adult Data	232
B.2 Bivariate Tables for The Child Data	235
B.3 Summary of Bivariate Statistical Tests	237
Appendix C Proportional Odds Assumption	238
Appendix D Codebook	241

List of Tables

4.1	Descriptive Statistics of The Adult Data (2002/03–2014/15 NZHS)	28
4.2	Descriptive Statistics of The Child Data (2006/07–2014/15 NZHS)	29
4.3	Missing Data Table	29
4.4	The Regression Summary of The Adult Data	36
4.5	The Regression Summary of The Child Data	37
B.1	Mean Age by BMI Tertiles	232
B.2	Sex by BMI Tertiles	232
B.3	Ethnicities by BMI Tertiles	232
B.4	Deprivation Quintiles by BMI Tertiles	233
B.5	Urban/Rural Area by BMI Tertiles	233
B.6	Household Income by BMI Tertiles	233
B.7	Educational Qualifications by BMI Tertiles	233
B.8	Adherence to Fruit Guideline by BMI Tertiles	233
B.9	Adherence to Vegetable Guideline by BMI Tertiles	233
B.10	Migration Status by BMI Tertiles	234
B.11	Difficulty Climbing Stairs by BMI Tertiles	234
B.12	Smoking Status by BMI Tertiles	234
B.13	Drinking Problem by BMI Tertiles	234
B.14	Physical Activity by BMI Tertiles	234
B.15	Sedentary Lifestyle by BMI Tertiles	234
B.16	Mean Age by BMI Categories	235
B.17	Sex by BMI Categories	235
B.18	Ethnicities by BMI Categories	235
B.19	Deprivation Quintiles by BMI Categories	235
B.20	Urban/Rural Area by BMI Categories	235
B.21	Household Income by BMI Categories	236
B.22	Parents' Educational Qualifications by BMI Categories	236
B.23	Adherence to Fruit Guideline by BMI Categories	236

B.24 Adherence to Vegetable Guideline by BMI Categories	236
B.25 Soft Drink Consumption by BMI Categories	236
B.26 Fast Food Consumption by BMI Categories	236
B.27 Bivariate Analysis of The Covariates and BMI Tertiles (Adult Data)	237
B.28 Bivariate Analysis of The Covariates and BMI Categories (Child Data)	237
C.1 Odds Ratios from Binomial Regressions with Contrasts in The Adult Data	239
C.2 Odds Ratios from Binomial Regressions with Contrasts in The Child Data	240
D.1 Codebook for Derived and Renamed Variables in This Thesis	241

List of Figures

2.1	Pathophysiological Pathway of Obesity-Associated Chronic Conditions	4
3.1	Theoretical Framework	19
4.1	BMI Categories in The Adult and Child Data	27
4.2	Bivariate Charts for The Adult Data	31
4.3	Bivariate Charts for The Child Data	32
4.4	Age-Standardised Obesity Prevalence by Ethnicity from 2002/03 to 2014/15	33
4.5	Age-Standardised Overweight/Obesity Prevalence by Ethnicity from 2002/03 to 2014/15	34

Listings

- A.1 Cleaning The NZHSs Data 64
- A.2 The NZHSs Data Analyses 178
- A.3 Bivariate Analyses - BMI 208
- A.4 Bivariate Analyses - Deprivation Quintiles 227

Acknowledgements

I wish to express sincere gratitude to Dr. Arindam Basu and Prof. Ray Kirk for their assistance in the preparation of this manuscript. I enjoyed every bit of my research and writing process because of their kind support and guidance. Not only that, they have also taught me how to critically formulate my research arguments and what it means to do research. Though the process has not been easy, this unforgettable journey allows me to wholeheartedly appreciate both the challenges and the joy of doing research even more.

I am also grateful to Statistics New Zealand for their assistance in providing me with the New Zealand Health Survey data. They also had assisted me many times with the questions I had regarding the data.

I would also like to express special thanks to Pat Coope whose assistance in statistics has taught me to better understand the importance of statistical tests and its implication in interpreting my research findings. I also want to thank Daniel Gerhard and Elena Moltchanova for setting up statistics workshops and introducing me to Bayesian inference. The knowledge and experience I gained helped me see the strange and peculiar things around the world with better clarity and comprehension.

Special thanks go to Robyn Johnston, Chris Chen, Lim Su Hui, Wade Stent, Kate Reid, Gopal Panta, Uzma Irfan, Llyween Cooper, Helen Mataiti, Khoa Vo, Arun Blasi, Babajide Onademuren, Mark Owusu, Vikram Lakshminarayanan, Ashleigh Sandy Gozali, and Anne Whitfield. I am really thankful for them, listening to their stories and experiences never fail to cheer me up. Their practical advice on daily life also had helped me on multiple occasions when I had no clue what I should do. Though it may seem trivial, the moral support, the jokes, and the laugh are the things that keep me going and make me smile.

Finally, I must express my very profound gratitude to my parents, my brother, my best friends in Indonesia for providing me with unfailing support throughout my years of study. This accomplishment would not have been possible without them. Thank you.

Abstract

Background: In New Zealand 31% of adults and one in nine children aged 2–14 years were obese in 2014/15. According to the 2014/15 New Zealand health survey, people who lived in the most deprived areas were almost two times more likely to be obese than those who lived in the least deprived areas.

Objectives: To assess the association between the area level deprivation and overweight/obesity prevalence while controlling for demographic variables and health-related behaviours.

Method: The analyses were based on the health surveys conducted from 2002/03 to 2014/15 in New Zealand. This study assessed area level deprivation using the New Zealand Deprivation Index. Proportional odds logistic regression with sampling weights was used to answer the objective.

Results: Adults who lived in deprivation quintile five were more likely (OR 1.46, 95% CI 1.33–1.60) to be in the higher Body Mass Index (BMI) tertiles compared with those who lived in quintile one, and the association was stronger in children (OR 1.76, 95% CI 1.50–2.06). Pacific (OR 8.41, 95% CI 7.22–9.79) and Māori adults (OR 3.01, 95% CI 2.75–3.29) had a higher odds to be in the higher BMI tertiles compared with Europeans.

Conclusion: Based on the representative nationwide dataset, I found that higher deprivation quintiles were significantly associated with a higher likelihood of being in the higher BMI categories after controlling for demographic information and health behaviours. Public health interventions targeting obesity should take into account the deprivation level and ethnic groups' composition of the communities.

Glossary

agnostic A position/attitude without prior beliefs about the subject. 9

AHEAD Action for Health in Diabetes. A multi-centre, randomised controlled trial, designed to determine whether intentional weight loss reduces cardiovascular morbidity and mortality in overweight individuals with type 2 diabetes. 7

anastomotic rupture A break in the surgically co-joined structure of the body. In this research context, it relates to the intestines or stomach. 8

apoptosis A regulated cell suicide process, usually triggered when the cells are no longer needed or posing a threat to the organism. 9

AUDIT Alcohol Use Disorders Identification Test (AUDIT) is a 10-item screening tool developed by the WHO to assess alcohol consumption, drinking behaviors, and alcohol-related problems. A score of 8 or more is considered to indicate hazardous or harmful alcohol use. 22, 28

bariatric surgery Surgical procedures on obese people to reduce their weight. Examples of such procedures are: reducing the size of the stomach with gastric band, removing a portion of the stomach, re-routing the small intestines to a small stomach pouch, etc. 2, 8

BMI Body mass index (BMI) is a person's weight in kilograms divided by the square of height in metre. vii, xi, 1, 2, 4–6, 8, 9, 12, 13, 15, 16, 21, 23, 26–29, 34, 35, 39–45, 47

BMI-H BMI-Heritability is the heritability coefficient of BMI. A proportion of the variation in BMI that can be explained by genetics. 8

Bogalusa Heart Study The Bogalusa Heart Study is a biracial (65% white and 35% black) community-based, long-term investigation of the early natural history of CVD beginning in childhood. 5

CDC Centers for Disease Control and Prevention. 5

CNS Central nervous system. A part of the nervous system consisting of the brain and spinal cord. 2, 9

comorbidities The presence of one or more medical conditions associated with the primary disorder. In this research context, any medical conditions associated with obesity such as type II diabetes melitus, hypertension, dyslipidemia, etc.. 1

CVD Cardiovascular disease. xi, xii, 5

DALY Disability-Adjusted Life Years (DALY) is a measure of overall disease burden, expressed as the number of years lost due to disability or early death. 4

doubly-labelled water A method that measures metabolic rate using an uncommon isotope of water (deuterium and oxygen-18) excreted from the body. 13

FHR The Framingham Heart Study (FHS) is a large cohort study in the town of Framingham, Massachusetts. It has led to the identification of major CVD risk factors and many others. 4

gastrointestinal Relating to the stomach and intestines. 4, 8

GDP Gross Domestic Product (GDP) is a monetary measure of the market value of all final goods and services produced in a given period (usually yearly) in a country. Nominal GDP estimates are commonly used to determine the economic performance of a country and to make international comparisons. 8

GNI Gross national income (GNI) is the total domestic and foreign output claimed by residents of a country. It consists of gross domestic product plus factor incomes earned by foreign residents, minus income earned in the domestic economy by non-residents. 11

GWAS Genome-Wide Association Study (GWAS) is an examination of many common genetic variants in different individuals to see if any variant is associated with a trait. 2, 9

HDL High-density lipoprotein (HDL) is one of the five major groups of lipoprotein. Increasing concentrations of HDL particles are strongly associated with decreasing accumulation of atherosclerosis within the walls of arteries. 5, 7

hemorrhage An escape of blood from a ruptured blood vessel. 8

HHID Household ID. It is a household identifier for the respondents of the NZHS. 24

HR Hazard ratio (HR) is the probability of an event (death or disease) occurring in the study group compared to the probability in the comparator group at any time point. 4, 5

IOTF International Obesity Task Force (IOTF) is an organisation designed to combat obesity around the world. Later known as World Obesity Federation. xiii, 1, 5, 21, 23

LDL Low-density lipoprotein (LDL) is one of the five major groups of lipoprotein and a risk marker for cardiovascular disease. Higher value of LDL translates to higher risk of cardiovascular disease. 5

NHANES The National Health and Nutrition Examination Survey (NHANES) is a survey that is designed to assess the health and nutritional status of adults and children in the United States. 7

nicotinic cholinergic receptors They are receptors that respond to the neurotransmitter acetylcholine. 16

NZHS New Zealand Health Survey. 1, 3, 15, 19, 20, 23–26, 45, 46

obstructive sleep apnea A medical condition where the breathing stop and start repeatedly during sleep caused by an obstruction in the airway. 8

OECD The Organisation for Economic Co-operation and Development (OECD) is an organisation that aims to promote policies that will improve the economic and social well-being of people around the world. 1, 40, 41

price elasticity A measure of the effect of a price change or a change in the quantity supplied on the demand for a product or service. 12

SES Socio-Economic Status. 3

SNP A Single-nucleotide polymorphism (SNP) is a variation in a single nucleotide that occurs at a specific position in the genome. For example, a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain location of DNA. 2, 9

SSB Sugar-sweetened beverage. 13

triglyceride It is an ester derived from glycerol and three fatty acids. A high concentration of triglyceride is associated with an increased risk of heart diseases, diabetes, and stroke. 5, 7

WC Waist Circumference. 1, 5

WHO World Health Organization. 5, 22, 23

WHR Waist/hip ratio. 1, 5

World Obesity Federation It is an organisation which has a mission is to lead and drive global efforts to reduce, prevent and treat obesity. Formerly known as IOTF. xii, 5, 21

1 Background

Obesity is a chronic condition associated with an increased risk of hypertension, dyslipidaemia, type II diabetes, stroke, congestive heart failure, coronary artery disease, increasing health costs, and many other comorbidities [Guh et al., 2009, Jensen et al., 2013, Barness et al., 2007]. Those chronic complications require resources to treat and in turn has led to the growing healthcare expenditure in the population. An individual with a body mass index (BMI) between 30 and 35 is estimated to require 19% more health expenses compared with those with normal BMI, and as the BMI breaks through 35 they would need 51% more [Buchmueller and Johar, 2015]. The rate of obesity around the world has doubled since 1980; with 39% and 13% of adults aged 18 and over, were overweight and obese respectively in 2014 [World Health Organization, 2015]. Among The Organisation for Economic Co-operation and Development (OECD) countries, New Zealand was ranked third in its adult obesity prevalence in 2017 [OECD, 2017]. According to the New Zealand Health Survey (NZHS) in 2014/15, 31% of adults and one in nine children aged 2–14 years were obese [Ministry of Health, 2015a]; the rates were higher among Māori, Pacific, and those with the highest level of socio-economic deprivation [Ministry of Health, 2015a, Swinburn et al., 2014]. From New Zealand's economic perspective, the estimated direct cost of treating obesity is NZ\$460 million in 2004, while the indirect cost is NZ\$370 million in the same year [Ministry of Health, 2009]. The actual direct cost calculated in 2006 was NZ\$624 million or 4.4% of New Zealand's total health care expenditure, in addition to NZ\$225 million due to lost productivity [Swinburn et al., 2014].

The criteria for overweight in adults is defined by a BMI between 25 and 30, and obesity is defined by a BMI of 30 or over [Centers for Disease Control and Prevention, 2016]. Whereas the criteria for defining overweight and obesity for those aged between two and below 18 years are based on the International Obesity Task Force (IOTF) BMI-for-age chart. It is defined in such a way so that it will merge with the adult's BMI criteria once they reach 18 years of age [Cole and Lobstein, 2012]. Other adiposity measurements such as waist circumference (WC) and waist/hip ratio (WHR) were thought to have superiority in predicting health risks associated with obesity. Nevertheless, a meta-analysis comparing BMI, WC, and WHR in predicting the incidence of diabetes found no significant difference [Vazquez et al., 2007]. For one standard deviation increase in BMI, WC, and WHR; the age-adjusted odds ratio for diabetes in men (women) were 1.52 (1.59), 1.54 (1.70), and 1.53 (1.50). For hypertension, they were 1.68 (1.55), 1.66 (1.51), and 1.45 (1.28) [Huxley et al., 2010]. Considering the high correlations across adiposity measurements and none showing superior discriminatory capability, it is clinically irrelevant to choose one measurement over the other [Vazquez et al., 2007, Huxley et al., 2010].

Individual management options for treating obese patients range from lifestyle interventions to invasive surgi-

cal procedures. Lifestyle interventions aim to improve physical activity levels and dietary habits through various training regiments and diet restrictions [Jensen et al., 2013]. However, such interventions only produce slight weight loss [Shaw et al., 2006], have a poor long-term adherence rate [Sheperd, 2005, Kovacs et al., 2014] and a high rate of regaining weight loss after stopping the intervention [de Vos et al., 2016]. On the other hand, bariatric surgery was able to achieve a long lasting weight loss and a significant reduction in cardiovascular events compared with those receiving non-invasive treatments [Sjöström et al., 2012]. Furthermore, there is a high rate of resolution of chronic conditions such as diabetes, hypertension and obstructive sleep apnea following bariatric surgery. Nevertheless, bariatric surgery is not only expensive but also has an operative mortality rate of 0.1–1.1% [Buchwald et al., 2004]. In addition, only those who are morbidly obese are eligible to undergo the surgery [Neff et al., 2013, Sheperd, 2005, Jull et al., 2011]. With the current available individual interventions, it does not appear as though these will be sufficient to curb the obesity epidemic at a population level.

Since the conception and increasing use of Genome-Wide Association Studies (GWAS), more than 20 loci have been found to have an association with obesity. Several of these genes are expressed in the Central Nervous System (CNS) which suggests involvement in energy homeostasis through feeding regulation [Herrera and Lindgren, 2010]. Nevertheless, two commonly identified variants of Single-Nucleotide Polymorphisms (SNPs) that have been linked to obesity, in an intron of FTO gene (situated in chromosome 16) and within a presumed regularly site 188kb downstream of the MC4R gene (situated in chromosome 18), only account for less than 2% variance in adult BMI. GWAS on obesity genetics may be able to shed more insights on the pathophysiology of obesity. Unfortunately, it will not be feasible anytime soon to find the genetic solution to tackle the current obesity epidemic at a population level [Bogardus, 2009].

Urbanisation has changed how people live in relation to their social as well as physical environment [Malik et al., 2013b]. The compact urban environment draws food outlets that provide cheap and unhealthy meals closer to homes [Kearney, 2010]. Advancement in technology has also increased the demand for mechanised works and activities, which in turn decrease the physical activity levels of a population. On top of that, people spend more of their leisure time watching television or browsing the Internet [Malik et al., 2013b]. Changes in agricultural practice in the past 50 years have also made our food supply bountiful and cheap [Kearney, 2010, Johnson and Wardle, 2011]. The world has increased its food consumption by 400 kcal per person per day between 1969/1971 and 2005/2007. Types of diets also shifted from staples (e.g., roots and tubers) into livestock products and vegetable oils, which are denser in energy [Alexandratos and Bruinsma, 2012, Kearney, 2010]. Food scarcity is no longer an issue in high-income countries with a strong welfare system [Dinsa et al., 2012]. Instead, the problem has shifted to an unhealthy living environment that promotes positive energy balance. Positive energy balance, either surplus of energy intake (e.g. energy-dense diet) or low energy output (e.g. sedentary lifestyle), can contribute to obesity [Giskes et al., 2011, Malik et al., 2013b, Kelly and Swinburn, 2015].

Disparities in the incidence of obesity exist across people with different socio-economic statuses (SES). Dinsa, et al. (2012) conducted a systematic review looking at the association between socio-economic status and obesity in low- and middle-income countries (defined by the World Bank as countries with an income per capita up to US\$12,275). Obesity is prevalent among high SES groups in low-income countries. Poor families in low-income countries still experience food scarcity and engage in manual labour; those factors protect them from obesity. Whereas, richer families tend to be overweight/obese because they have access to surplus food [Dinsa et al., 2012] and excess weight can be perceived as a sign of wealth [Wu et al., 2015]. As the income level of a country increases, the relationship between SES and obesity becomes less clear. This may be explained by less apparent food shortages even among poor families in middle-income countries as these countries have more resources to help the poor [Wu et al., 2015, Dinsa et al., 2012]. In contrast, low SES is associated with higher risk of overweight/obesity in high-income countries [Wu et al., 2015, Newton et al., 2017]. Poor families in these countries consume low-cost energy-dense food and have less opportunities to participate in physical activities while their well-off counterparts have more resources to choose healthier diet and lifestyle [Wu et al., 2015]. These tell us that the association between SES and obesity is not straight-forward, and the direction of the association changes depending on the country's income level.

In New Zealand, according to the 2014/15 health survey, people who lived in the most deprived areas were twice as likely to be obese than those who lived in the least deprived areas [Ministry of Health, 2015a]. Other studies in the United States (US) [Eagle et al., 2012], the United Kingdom (UK) [Stafford et al., 2010], and Sweden [Li et al., 2014] also support the finding that people who live in a deprived neighbourhood are more likely to be obese. In addition, the incidence of type II diabetes is higher in more deprived areas [Evans et al., 2000, Maier et al., 2013]. These suggest there may be underlying processes of socio-economic indicators affecting the incidence of obesity and diabetes. To my knowledge, there is no study in New Zealand that assesses the association between socio-economic deprivation and obesity. This study will assess the association between area level deprivation and overweight/obesity prevalence using individual data from the NZHS in 2002/03 to 2014/15. Relevant demographic information and behaviours that may influence obesity prevalence will be controlled.

2 Literature Review

2.1 Health Risks Associated with Obesity

Obesity is a growing problem in New Zealand. Not only it is a major contributor to Disability-Adjusted Life Years (DALY) lost but obese people also are associated with an increased risk of developing chronic conditions [Ministry of Health, 2009]. Chronic conditions such as hypertension and diabetes mellitus type II require long term management and once developed, they are irreversible and require a lot of resources to manage. Additionally, obesity is associated with a higher risk of cardiovascular diseases, pre-eclampsia, respiratory problems, depression, infertility, gastrointestinal diseases, cancers, and a lower life expectancy [Banjare and Bhalerao, 2016]. From a biological perspective, obesity is known to cause insulin resistance, endothelial dysfunction, and promote atherosclerosis formation. All of these lead to the development of chronic non-communicable diseases typically found in obese people [Antonini et al., 2007]. The pathophysiology of obesity-associated chronic conditions is outlined in Figure 2.1.

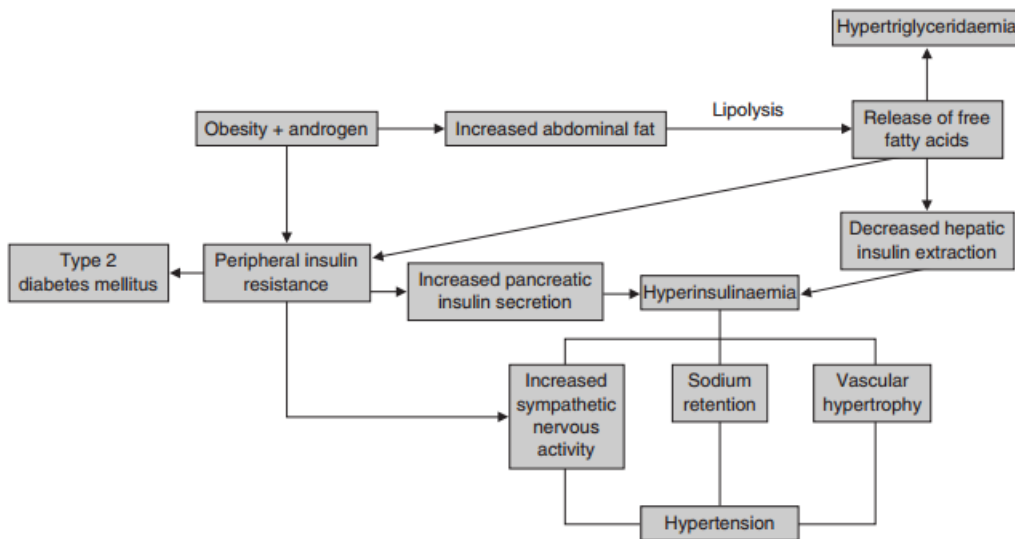


Figure 2.1: Pathophysiological Pathway of Obesity-Associated Chronic Conditions [Antonini et al., 2007]

The risk of obesity-associated chronic diseases is thought to positively associate with the duration of obesity. A study using the data from the Frammingham Heart Study (FHR) have found that obesity duration was associated with the risk of developing type II diabetes. the Hazard Ratio (HR) for type II diabetes mellitus was 1.11 for men and 1.06 for women per additional 2-year increase in the duration of obesity [Abdullah et al., 2011]. However, the association between obesity duration and type II diabetes was heavily attenuated when time-varying BMI

and current BMI were included in the analysis. The HR for type II diabetes after the adjustment was 1.02 per additional 2-year increase in the duration of obesity [Hu et al., 2014]. Even though both current BMI and duration of obesity are significant, current BMI appears to exert a stronger influence in predicting the risk of developing type II diabetes.

The health risk associated with obesity in early age is carried to adulthood. Obese children and adolescents have higher fasting glucose, fasting insulin, triglyceride, systolic blood pressure, prevalence of impaired glucose tolerance and lower high-density lipoprotein (HDL) cholesterol compared with their non-obese counterparts [Weiss and Kaufman, 2008]. The Bogalusa Heart Study found that a higher BMI and blood pressure in childhood were associated with higher left ventricular mass when they reached adulthood [Lai et al., 2014]. Multiple prospective cohort studies in Finland also asserts this finding [Juonala et al., 2011]. Adults who had always been obese since childhood had an increased risk of type II diabetes and an elevated levels of cardiovascular (CVD) risk biomarkers (i.e. LDL, HDL, triglyceride, and intima-media thickness of carotid artery) compared with adults who always had a normal BMI. Moreover, non-obese adults who were overweight or obese during childhood had similar risk to those who had normal BMI from childhood to adulthood [Juonala et al., 2011]. Even though it is hard to reverse childhood obesity (80% of obese children are reported to become obese adults [Cali and Caprio, 2008]), successfully reversing obesity in children will reduce the risk of developing type II diabetes and CVD later in life. This further put an emphasis on the importance of preventing obesity earlier in life.

2.2 Defining Overweight and Obesity

There are various criteria used in defining childhood overweight and obesity. The Centers for Disease Control and Prevention (CDC) growth chart uses 85th percentile and 95th percentile to define overweight and obesity respectively [Centers for Disease Control and Prevention, 2013]. However, the cut-offs used in the CDC and World Health Organization (WHO) growth chart are designed to represent the US population whom always have had a higher average BMI compared with other countries around the world [Pedrosa et al., 2011, Cole and Lobstein, 2012]. Whereas, the World Obesity Federation (formerly known as the International Obesity Task Force (IOTF)) designs the cut-off points using survey data on children aged 6–18 years from multiple countries (Brazil, United Kingdom, Hong Kong, Singapore, Netherlands, and USA), which are more representative of the world population compared to the CDC. The overweight criteria according to the IOTF cut-off is close to 90th percentile while the obesity cut-off is above the 98th centile [Cole and Lobstein, 2012].

Various anthropometric measurements to define overweight obesity such as BMI, WC, and WHR have a lot of limitations. Even though they have variable strength in predicting adverse health events, a meta-analysis confirmed that no measurement is superior to the others [Vazquez et al., 2007]. It is clear that there is a continuous

positive association of increasing weight with the risk of adverse health outcomes. Nevertheless, there is no BMI thresholds that could divide the population into meaningful risk groups [Stommel and Schoenborn, 2010]. In general, the relationship between BMI and all-cause mortality follows a U-shaped pattern. The risk progressively increases as it approach towards both ends of the spectrum. However, the relationship turns into a J-shape among healthy individuals (i.e. the risk of adverse health outcomes in the low BMI group is higher than those in the middle BMI group, but not as high as those in the high BMI group). It was later found that people in the lower BMI spectrum were confounded with their smoking habit, which explains why people in the lower BMI spectrum have a worse outcome [Antonopoulos et al., 2016]. Age is also known confound the relationship between BMI and mortality risk. Being overweight in old age is associated with lower mortality risk compared with those in the normal BMI category [Flegal et al., 2013]. Furthermore, compared with normal-weight group, mortality rate in the overweight and mildly obese group with stable coronary heart disease is lower even after adjusting for confounding factors [Romero-Corral et al., 2006]. Using BMI as a sole indicator in predicting adverse health outcomes without controlling for other factors such as age and health-related behaviours can produce erroneous predictions.

A lot of problems also arise when using universal BMI cut-offs to define overweight/obesity in diverse ethnic groups because body weight distribution and predisposition to store visceral fat tissues differ across ethnic groups. For example, African-American and Hispanic people tend to have the highest average BMI and Asian people have the lowest [Stommel and Schoenborn, 2010, El-Sayed et al., 2011]. In terms of body fat percentage, Chinese, Malaysians, Indians, Indonesian and Japanese people have a higher body fat percentage at a low BMI [Barba et al., 2004]. Also, Māori, Pacific and Indian people had a higher body fat composition compared with the Europeans at the same BMI level. A BMI of 30 kg/m² in European men (women) would have an equal body fat percentage as a BMI of 31 (33), 34 (35), and 24 (26) in Māori, Pacific, and Asian Indian respectively [Rush et al., 2009]. Higher body fat percentage has been linked to a higher risk of developing type II diabetes and cardiovascular disease in the Asian population compared with European people of the same age, sex, and BMI [Barba et al., 2004]. A large cross-sectional study in the US also found that the prevalence of chronic health conditions such as hypertension, diabetes, coronary heart disease, asthma, and arthritis varies across ethnic groups regardless of BMI levels. Having said that, after adjusting for socio-demographic and behavioural factors, there are no clear BMI thresholds that divide the population into meaningful risk groups. The only consistent finding is that increasing BMI levels are associated with higher rates of chronic diseases across all ethnic groups [Stommel and Schoenborn, 2010]. It may be better to use ethnic-specific BMI standard deviation when assessing the relationship between BMI and obesity-associated health risk among diverse ethnic groups.

2.3 Lifestyle Interventions

The discourse regarding the cause of the obesity epidemic has been framed on bad dietary habits and low levels of physical activity, suggesting that the responsibility lies on the individuals and can be modified through education as well as physical exercise [Luke and Cooper, 2013]. The Look AHEAD trial, an 8-year multi-centered randomised trial, compared an intensive lifestyle intervention with a less-intensive diabetes support and education on weight reduction. The intervention group achieved maximum weight loss (8.5% of initial weight) in the first year, only to regain their weight back at 4.0%–4.7% loss for the remainder of the trial. Conversely, the comparison group managed to achieve 2.1% weight loss at the end of study [Wadden et al., 2014]. The small differences of weight-loss achieved between intensive lifestyle intervention and usual diabetes support was not substantive enough to make a difference. On top of that, a meta-analysis [Shaw et al., 2006] and long-term clinical trials [de Vos et al., 2016, Wadden et al., 2014] assessing the effect of exercise on overweight and obesity found that most people lost weight initially but regain their weight loss over time, further casting doubt on the long-term effect of exercise and diet-related intervention on maintaining weight loss.

Another problem with lifestyle interventions is their poor compliance rate [Kovacs et al., 2014]. Despite the majority of overweight and obese people wanting to reverse their conditions, expecting them to exercise more also appears to be unrealistic. The National Health and Nutrition Examination Survey (NHANES) measured the duration and intensity of physical activity using an accelerometer, and it found that only 0.3% of adults achieved the current physical activity guideline which recommends 150 minutes of moderate activity or 75 minutes of vigorous activity per week [Luke et al., 2011]. Moreover, some people with a limited functional capacity due to medical conditions are more prone to becoming obese and unable to meet the required physical activity level [Banjare and Bhalerao, 2016]. Considering only a small proportion of the population met the physical activity guideline and the possibility that overweight/obese people would have more difficulty in achieving it, it is very unlikely that increasing population physical activity level would have an impact on obesity prevalence [Luke and Cooper, 2013].

Lifestyle interventions, however, provide health benefits regardless of weight loss [Shaw et al., 2006, Luke and Cooper, 2013]. The results of the Look AHEAD study at four years found that those who were in the intensive lifestyle intervention group achieved greater improvements in blood pressure, triglyceride, and HDL level [Wing et al., 2010]. Furthermore, a 20-year study in China among adults with impaired glucose tolerance found that those who received lifestyle intervention for 6 years had 43% lower diabetes incidence than the control group. Albeit, at the end of 20-year mark, 80% of people in the intervention group had diabetes and the hazard risk ratio of cardiovascular events was inconclusive between the two groups [Li et al., 2008]. In general, a combination of physical exercise and diet is better than either one of them alone in improving markers of cardiovascular disease and diabetes regardless of weight change [Shaw et al., 2006]. Therefore, there is no benefit in questioning the

public health value of maintaining a high level of physical activity and healthy diets, and they should be addressed in conjunction with other public health strategies targeting different aspects of health.

Another option for people wanting to reduce their weight is through the pharmacological approach. A meta-analysis on commonly used anti-obesity drugs (Orlistat, Sibutramine and Rimonabant) had been proven to be moderately effective in achieving weight loss. On average, those who used the drugs achieved 5 kg or less of placebo-subtracted weight reductions. Nonetheless, the major barrier to anti-obesity drugs was the lack of adherence, with an average of 30% to 40% attrition rate. The low adherence rate can be explained by potential side effects associated with the use of anti-obesity drugs such as gastrointestinal symptoms, increased blood pressure, and mood disorders [Padwal et al., 2003].

The more invasive strategy for weight loss is through bariatric surgery. In terms of efficacy, the mean ten-year weight reduction ranged from 13% to 25% [Neff et al., 2013]; also, 77% of diabetes, 86% of obstructive sleep apnea, and 62% of hypertension were resolved following the procedure [Buchwald et al., 2004]. A prospective controlled study in Sweden found that the bariatric surgery group had an adjusted hazard risk of 0.67 for cardiovascular events compared with the obese adults receiving usual care (i.e. ranging from lifestyle interventions at many sites to no intervention in other sites) [Sjöström et al., 2012]. Despite its efficacy, bariatric surgery is related to surgical complications such as bowel perforation, bowel obstruction, hemorrhage, anastomotic rupture, micro-nutrient deficiencies, and many others [Neff et al., 2013]. Combined with the fact that only those with a BMI of over 35 kg/m² are eligible for this procedure [Jull et al., 2011, Jensen et al., 2013, Sheperd, 2005, Neff et al., 2013] and an operative mortality rate of 0.1–1.1%, making it unlikely to resolve the obesity problem at a population level.

2.4 The Genetics of Obesity

Studies of monozygotic and dizygotic twins have shown that body weight is highly heritable, which means a proportion of the inter-individual difference in body weight can be explained by genetic influence [O’Rahilly and Farooqi, 2008, Herrera and Lindgren, 2010, Barness et al., 2007]. The coefficient of BMI heritability (BMI-H) is found to vary from 31% to 90% depending on the characteristics of the population in the study [Min et al., 2013]. For example, BMI-H is influenced by the average GDP and average BMI of the population. A population with higher average BMI and higher GDP is more likely to have a greater BMI-H. Additionally, BMI-H is also influenced by age. Unique environments are relatively more diverse in adulthood compared with childhood; thus, BMI-H is lower in older population [Min et al., 2013]. This variation happens because the heritability equation assumes that genetic and environmental factors are independent of each other. In reality, we cannot untangle the complex interaction of the gene-environmental relationship. For example, consider a scenario about the

behaviour of parents and their children. Genetics may influence the parents' behaviour which in turn influence the environment of the child (e.g. intelligent parents raise their children in an intellectual environment) or the behaviour of the child, which is influenced by genetics, affect their parent's response which is part of the child's environment (e.g. parents respond with frustration to an irritable child) [Stenberg, 2013].

GWAS is designed to detect genetic variants in the genomic loci that are associated with observable characteristics of the population. For example, it can detect associations between common SNPs and common diseases such as heart disease and diabetes [Visscher et al., 2012]. This breakthrough genetic analysis has allowed researchers to capture about 80% of all common variants using as few as 500,000 chosen SNPs. This translates to 20 loci being identified as having an association with body weight [Herrera and Lindgren, 2010]. Among those loci, FTO has been repeatedly identified. Some genes in the FTO region are highly expressed in the brain and play a role in appetite regulations [Mei et al., 2012, Herrera and Lindgren, 2010]. Similar to FTO, other single gene mutations are also shown to mediate the development of obesity through appetite regulating CNS mechanisms [Herrera and Lindgren, 2010, O'Rahilly and Farooqi, 2008], while the rest are known to be involved in fat cell apoptosis and proliferation [Mei et al., 2012].

Despite the success of the GWAS strategy, the established loci only explains around 2% of inter-individual BMI variation [Herrera and Lindgren, 2010, Mei et al., 2012, Bogardus, 2009, Min et al., 2013]. There is a discrepancy in the heritability estimates from the twin studies (31–90%) [Min et al., 2013] and the GWAS approach (around 2%) [Bogardus, 2009], which is called the missing heritability. Although GWASs are agnostic with respect to prior beliefs, they are biased in terms of what is detectable. GWASs are only able to detect causal variants that are common in the population and assume that common diseases are likely to be caused by common variants. However, we cannot dismiss the possibility that rare variants might have a large effect on the population characteristics and they would have been missed by the GWAS approach [Visscher et al., 2012]. Even though GWAS is able to explain the pathophysiology of diseases, the practical utility is questionable. In particular, identifying all 18 types of II diabetes susceptible genes are no better in predicting diabetes than relying only on the person's BMI, age, and sex [Bogardus, 2009]. Further casting doubt on its practicability in predicting disease risk when easier to assess information can make the same prediction.

2.5 Agriculture

The changes in agricultural practice over the past 50 years has provided our world with a greater diversity of foods and less seasonal dependence, thereby changing our dietary practice tremendously. There are two stages of change in the pattern of food consumption, the first one is the "expansion stage", which is followed by the "substitution stage". The expansion stage is characterised by the effort to increase total caloric intake through cheaper

foodstuffs rich in carbohydrate. Whereas, in the substitution stage, the total calories consumed do not differ much but there are shifts from carbohydrate-rich staples to vegetable oils and animal products [Kearney, 2010]. These shifts can be seen from the pattern of world food consumption in the past decades. In 1969/71, people around the world on average consumed 2370 kcal/day and the number increased to 2770 kcal/day in 2005/07. The increase was more prominent in low-middle-income countries, despite some parts within the countries still experiencing a food shortage and a high incidence of undernourishment due to unequal food distribution [Alexandratos and Bruinsma, 2012]. Vandevijvere et al. (2015) assessed the relationship between national average food energy supply and body weight using survey data from 56 countries. They found that an increase in food energy supply is significantly associated with an increase in average population weight among high-income countries but not for low- and middle-income country groups. Nevertheless, as food becomes more abundant and a potential decrease in physical activity levels in low- and middle-income countries, the same association between energy intake and body weight is expected to occur in the future [Vandevijvere et al., 2015].

2.6 Urbanisation

Urbanisation may also play a role in shaping the environment that promote the development of obesity. Urbanisation changes how people get their food, where lax trade policies that draw multinational food companies to invest increase the availability of food outlets. As food outlets grow closer to homes, cheap and unhealthy foods become more readily available [Kearney, 2010]. More people in one area also mean that there is a need for a better transportation system, increasing the availability of roads and highways. This, in turn, increases the use of private motorised transportation which reduces physical activities [Malik et al., 2013b, Sturm and An, 2014, Kearney, 2010]. Additionally, technological advancement reduces the need of manual labours as these tasks become more commonly performed by machines and computers. Leisure time is also spent on more sedentary activities such as television viewing, Internet use, and computer gaming [Kearney, 2010, Malik et al., 2013b, Dugas et al., 2011].

In the US, according to a meta-analysis, rural children are 26% more likely to be obese compared with urban children, despite obese rural children being more physically active [Johnson and Johnson, 2015]. Similar findings are also observed in China. Contrary to the stereotype of a higher physical activity level in rural areas, people in rural China tend to have a lower physical activity level due to agricultural mechanisation [Tian et al., 2014]. It was thought that rural dwellers have poorer diet due to lack of knowledge and lower level of education. However, a recent study found that urban dwellers consume more animal products rich in fat while having the same total caloric intake than their rural counterparts [Zhang et al., 2017]. This suggests that diets with higher total caloric intake may have a stronger influence than the protective effect of being physically active on obesity.

In contrast, studies in India [Ebrahim et al., 2010] and South-East Asian countries [Angkurawaranon et al., 2014] found that obesity is more prevalent in urban areas. Not only rural men in India had lower obesity prevalence; but they also had lower blood pressure, lipids, and fasting glucose than urban men [Ebrahim et al., 2010]. It seems that the strength of this association is also influenced by the gross national income (GNI) per capita of the countries. Countries with lower GNI per capita, like Vietnam and Laos, had a stronger association; people who lived in urban areas were three times more likely to be obese than those who lived in rural areas. Whereas, rural dwellers were 29% more likely to be obese compared with urban dwellers in Malaysia and Philippines (i.e. countries with higher GNI per capita) [Angkurawaranon et al., 2014]. Surprisingly, despite New Zealand being a high-income country, the association between urban/rural areas and obesity is similar to countries with lower income. In 2002, according to the Children's Nutrition Survey in New Zealand, urban boys were 1.3 times more likely to be obese than those who lived in rural areas, and girls were 1.4 times more likely. There was also no difference in total energy intake per day and the frequency of bouts of physical activity between the two groups [Hodgkin et al., 2010]. These show that there may be underlying mechanisms, other than total energy intake and physical activity, which contribute to the disparity in obesity prevalence between rural and urban areas.

2.7 Food Environment

There have been many studies assessing the role of the environment in shaping the obesity epidemic in recent years. Many environmental studies have looked at how the built environment might affect the levels of physical activity and dietary habits. The current living environment is thought to increase the availability of unhealthy food products (e.g. high-fat foods and sugar-sweetened beverages) and remove spaces for physical activities (e.g. increase private motorised transports) [Giskes et al., 2011]. Despite the sound arguments of environmental theories in trying to explain the obesity epidemic, most of the evidence from epidemiological research has been contradictory [Giskes et al., 2011, Durand et al., 2011, Mattes and Foster, 2014, Sturm and An, 2014].

Studies on accessibility factors (e.g. locations of supermarkets, food store density, availability of healthy foods, etc.) have received considerable attention lately. It was presumed that higher food store density will affect a person's body weight and increase the obesity prevalence in the area [Giskes et al., 2011]. A cross-sectional study in Canada, examining children's weight status and their neighbourhoods' characteristics, found that children had a lower risk of being overweight or obese when they had greater access to more affordable healthy food options within walking distance from their homes [Le et al., 2016]. However, a longitudinal study in the US found no significant impact of food outlet density on children's body weight [Chaparro et al., 2014]. Overall, the relationship between environmental factors and dietary intake have shown that accessibility to supermarkets and the availability of fruits and vegetables in the shops were not associated with fruit and vegetable consumption.

The conflicting evidence emerged due to difficulties and challenges in designing environmental studies. There is no standardised method of assessing the environmental features such as distance to stores and what type of foods were sold in the shops [Giskes et al., 2011]. In addition, assessment on accessibility to certain food shops relies heavily on the assumption that people would buy foods from nearby stores [Le et al., 2016]; discounting the idea that people might travel further to buy their foods.

2.8 Food Price and Consumption

Another important element of the food environment is price. Food price is thought to affect consumption which in turn will influence body weight. Two cohort studies in children have found that one standard deviation increase in fruit and vegetable price is associated with a 4% decrease in consumption [Sturm and Datar, 2011] and a 0.09 increase in a child's BMI z score [Morrissey et al., 2014]. A similar finding is also observed in young adults with an own-price elasticity estimate of -0.32, which is inelastic. However, the elasticity estimate became -0.66 among low-income young adults, indicating that fruit and vegetable consumption in this group is more sensitive to price change [Powell et al., 2009]. Even though the effect of price changes are small, they would have more influence on younger population from a lower income groups.

The association between soft drink/fast food price and food consumption is even less clear. A cohort study in the US found that there is no significant association between fast food or soft drink prices and their consumption among children [Sturm and Datar, 2011]. Moreover, another cohort study assessing the influence of soft drink and fast food prices on children's weight also found that they are not significant predictors for BMI [Morrissey et al., 2014]. The non-significant association between the changes in fast food prices and BMI is also observed among adults. This may have happened because the price of fast food is already low thus it was unlikely that price changes would have a meaningful influence on consumption. It was also possible that the reduction of fast food consumption due to an increase in price was substituted with an increase in consumption of other food products with the same caloric content, hence no weight change was observed [Cotti and Tefft, 2013].

In general, people perceived the cost of healthier food such as fruit and vegetables to be more expensive and this serves as a major barrier in acquiring them [Dijkstra et al., 2015, Powell et al., 2009, Ni Mhurchu and Ogra, 2007]. A study in the Netherlands has found that lower socio-economic groups adhered less often to healthy dietary guidelines and were more likely to perceived cost as a barrier to buying fruit, vegetables and fish [Dijkstra et al., 2015]. The same situation is also observed in New Zealand. When comparing the average price of regular baskets with the healthier ones showed that it was NZ\$36.27–NZ\$49.18 more expensive to have a healthier diet. It is not only more expensive, but also finding healthier food options is harder for people living in the less affluent areas. Imposing healthy dietary guidelines to these groups of people will only create frustration and will likely

fail in achieving the public health goals [Wang et al., 2010]. Having said that, informing the public about healthy diets such as avoiding saturated fats, reducing total caloric intake, and eating more fruits and vegetables will still be helpful in empowering the public to make better dietary choices. Although, it needs more than just dietary recommendations to halt the obesity epidemic [Woolf and Nestle, 2008].

2.9 Sugar-Sweetened Beverage and Fast Food Consumption on Body Weight

Soft drinks typically refer to beverages that contain carbonated water, but not all soft drinks contain a high level of sugar. A meta-analysis assessing the association between soft drink consumption and body weight found that studies which focused on Sugar-Sweetened Beverage (SSB) had greater effect on weight than those that used the term 'soft drink' [Vartanian et al., 2007]. Higher consumption of SSBs is consistently associated with higher body mass in both adults and children; and the effect is stronger in adults [Malik et al., 2013a, Vartanian et al., 2007]. Apparently, people who consumed SSBs do not compensate for the added energy and tend to increase their total food intake from other sources as well. Not only is sugar known to decrease satiety, but it is also thought to calibrate preference to sweet foods which can lead to an increased intake of sugar-rich food. SSB consumption is also associated with lower consumption of other nutrients such as calcium and vitamins [Vartanian et al., 2007]. On top of that, compared with individuals who consumed none or less than one serving of SSB per month, those who had 1–2 servings of SSB per day had a pooled risk of 1.26 (95%CI: 1.12–1.41) in developing type II diabetes [Malik et al., 2010].

Consumption of fast food, which often contains excessive amounts of refined starched and added sugars, is believed to contribute to obesity. A 15-year prospective study found that people who consumed more than two servings per week of fast food gained 4.5kg more and had two times greater increase in insulin resistance compared with those who consumed less than one serving per week [Pereira et al., 2005]. However, a study in China did not find any association between fast food consumption and BMI z-score in adolescents [Xue et al., 2016]. It appears that the effect of increased fast food consumption on higher BMI is consistent in adults, but less so in children and adolescents. It is possible that children are still growing and need more energy, thus extra caloric intake is not readily converted into excess body fat [Rosenheck, 2008].

2.10 Sedentary Environment

An environment that promotes sedentary lifestyle was also thought to increase the occurrence of obesity. It was assumed that people living in higher income countries would have lower physical activity and lower energy expenditure as they had more mechanised labour. However, a meta-analysis of doubly-labelled water studies found that energy expenditure did not differ according to the social and economic development of a country

despite significant differences in body size, further questioning the lack of physical activity as the primary driver of obesity [Dugas et al., 2011]. Using a model to estimate the change in energy intake that could theoretically contribute to the change in average body size showed that increases in food energy supply alone was sufficient to explain the growth of population weight [Vandevijvere et al., 2015]. A systematic review on built environment also found that the living environment that promotes physical activity (i.e. environment with more mixed land use, open spaces and walk paths) had no significant association with physical activity level or body mass [Durand et al., 2011]. However, most of these studies were cross-sectional and did not take into account neighbourhood self-selection (i.e. physically active people tend to live in an environment that supports their preferences). Even so, the evidence regarding the association between environment and physical activity are mixed even after accounting for neighbourhood self-selection and focusing on quasi experimental designs. Most associations found were in the expected direction or null, which imply that communities with mixed land use or open spaces may increase physical activity of the residents [McCormack and Shiell, 2011]. Despite that, more long-term longitudinal studies are needed to establish a causal link.

2.11 Socio-Economic Status

Research have shown that the socio-economic status is consistently associated with obesity status. Albeit, the direction of the association is dependent on the country's socio-economic level. In low-income countries, the more affluent families had greater overweight or obesity rates. Whereas in high-income countries, they had a lower risk of being overweight or obese [Fruhstorfer et al., 2016, Dinsa et al., 2012, Wu et al., 2015]. This happens because food availability remains a challenge in low-income countries, and only a portion of the communities have more than enough to eat and overweight is still perceived as a sign of wealth [Wu et al., 2015]. In contrast, high-income countries are able to secure enough food even for the lower socio-economic groups, especially those with a strong welfare system. Nevertheless, poor families in these countries do not have the resources to choose healthier diets and lifestyle. Whereas, the wealthier families are able to adhere to healthy diets, which tend to be more expensive [Dinsa et al., 2012, Sturm and An, 2014]. The same pattern of association is also observed in regards to the effect of education level on body weight. Higher education level is associated with a lower risk of obesity in high-income countries, while it is associated with a higher risk of obesity in lower-income countries [Cohen et al., 2013].

A different approach in measuring socio-economic status is by using the degree of the deprivation of an area. Area level deprivation is an aggregated index that measures multiple socio-economic indicators such as income, employment, education, housing, social services, etc. Many studies have shown that area level deprivation is associated with the disparities in obesity risk [Nau et al., 2015, Li et al., 2014, Stafford et al., 2010]

and various health outcomes [Maier et al., 2013, Singh et al., 2015, Exeter et al., 2015]. Children who lived in communities with higher socio-economic deprivation have a faster BMI growth compared with those who lived in less deprived areas [Nau et al., 2015]. Children in the deprived communities also had a poor quality diet [Craig et al., 2016], despite the mothers' knowledge about healthy diet being rated as high [Crombie et al., 2008]. The relationship between area level of deprivation and obesity persists even after adjusting for individual and family-socio economic status. This implies that there is a unique pathway in which area level deprivation influences the risk of being obese [Li et al., 2014]. When compared with people who lived in the least deprived areas, those in the most deprived areas also have higher rates of type II diabetes [Evans et al., 2000, Maier et al., 2013].

The Family Food Environment Survey of 136 New Zealand families found that people in the low-income group purchased fewer vegetables despite the fact that all income groups in the survey had basic kitchen amenities and did not have trouble accessing food shops. This shows that the majority of households in New Zealand would have less structural challenges (i.e. lack of facilities to prepare foods) in modifying their dietary habits [Smith et al., 2010]. Nonetheless, according to the NZHS 2014/15, people who lived in the most deprived areas were almost two times more likely to be obese than those who lived in the least deprived areas [Ministry of Health, 2015a]. Furthermore, the result from the NZHS 2015/16 showed that the rates of extreme obesity in people who lived in the least deprived areas have not changed since 2006/07, while people who lived in the most deprived areas experienced an increase in their extreme obesity and obesity rate during the same period [Ministry of Health, 2016].

There are possible explanations to why people who lived in an area with higher deprivation have higher rates of obesity and worse health outcomes. Socio-economically deprived communities may affect health outcomes through regional norms or attitudes towards health [Maier et al., 2013]. Behaviours that are detrimental to health such as smoking, alcohol abuse, and over-consumption of unhealthy energy-dense food may be more prevalent. Lower education level in the area may also result in less emphasis on nutrition and physical activity [Eagle et al., 2012, Nau et al., 2015]. On top of that, deprived communities tend to have higher psychosocial stresses, lower social capital, worse physical infrastructure and an inadequate access to health care services [Maier et al., 2013, Li et al., 2014, Nau et al., 2015, Cummins and Macintyre, 2006]. As opposed to wealthier communities, people in deprived areas would have more challenges in achieving a healthier lifestyle and accessing health care resources, thus they are more at risk of developing obesity and obesity-associated adverse conditions.

2.12 Smoking and Obesity

Multiple studies have found that current smokers tend to be leaner than those who never smoked or had quit smoking [Siahpush et al., 2014, MacKay et al., 2013]. Tobacco contains many harmful components such as

nicotine, carbon monoxide, and other stress oxidants. Nicotine binds to the nicotinic cholinergic receptors in the brain and activates the sympathetic nervous system, which then causes an increase in heart rate, blood pressure, and cardiac work. The increase in sympathetic activity also increases energy expenditure which explains the lower body weight observed among current smokers [Rigotti and Clair, 2013]. Age also plays an important role in this relationship. A study in Scotland found that people who had smoked more than 20 years were 50% more likely to be overweight than those who had never smoked. They also found that those aged 16–24 years who quit smoking were no more likely to be overweight or obese than current smokers or people who never smoked [MacKay et al., 2013]. Although current smokers tended to have normal weight, their overall mortality risk was higher than overweight or obese ex-smokers [Siahpush et al., 2014]. Smoking cessation among the younger population is desirable because it is not associated with weight gain and would reduce the mortality risk. Despite potential weight gain associated with smoking cessation among older population, it should not deter them from quitting considering an overall protective effect from mortality that they would gain.

2.13 Healthy Immigrant Effect

Many studies in the US, Canada, Australia, and the United Kingdom have found that immigrants have a better health status than the citizens of the recipient-countries [Vang et al., 2015, Kennedy et al., 2015, Goel et al., 2004]. In terms of obesity, a study in the US found that recent immigrants had significantly lower prevalence of obesity compared with the US-born citizens. However, the obesity rates of the immigrants began to climb up after 10 years and caught up with the host prevalence after 15 years of living in the US [Goel et al., 2004]. The protective effect of being an immigrant is less clear on children and is reversed for expecting mothers, older immigrants [Vang et al., 2015], and refugees [Norredam et al., 2014]. In general, the healthy immigrant effect can be explained by the immigration selection favoring highly educated and healthier immigrants [Kennedy et al., 2015]. Nonetheless, this selection process is less relevant in the perinatal period, childhood, adolescence, elderly, and refugees; which explains why the healthy migrant effect is not observed in these groups [Vang et al., 2015, Norredam et al., 2014].

2.14 Literature Review Summary

This chapter addresses that obesity is an important issue because it is associated with a higher risk of developing chronic conditions such as type II diabetes and CVDs. Defining overweight and obesity in the population is not easy, especially when dealing with diverse ethnic groups as they carry different risks at the same BMI level. It is also clear that tackling obesity at a population level through individual interventions are not enough, therefore, studying which environmental factors that have a strong influence on obesity is necessary to devise better public

health interventions. Multiple studies have found that people who lived in more deprived areas had a higher likelihood of becoming overweight/obese. It is important to study this relationship to reduce the health disparities across areas with different area level deprivations. This study will analyse this relationship while controlling for possible confounding factors which I have discussed in this chapter.

3 Population and Methods

3.1 Theoretical Framework

Obesity prevailed despite efforts to reduce obesity rates through lifestyle and pharmacological interventions. Interventions that try to improve physical activity levels and dietary habits fail to sustain long-term weight loss [Padwal et al., 2003, Shaw et al., 2006, Wadden et al., 2014]. Environmental factors play an important role in the obesity epidemic because they determine accessibilities to unhealthy/healthy food, open spaces for physical activity and health care services [Giskes et al., 2011, Sturm and An, 2014]. Areas with high deprivation level have social and structural challenges that put people at a higher risk of being overweight/obese [Nau et al., 2015, Li et al., 2014, Stafford et al., 2010]. The theory which serves as the foundation of this research is shown in Figure 3.1 below.

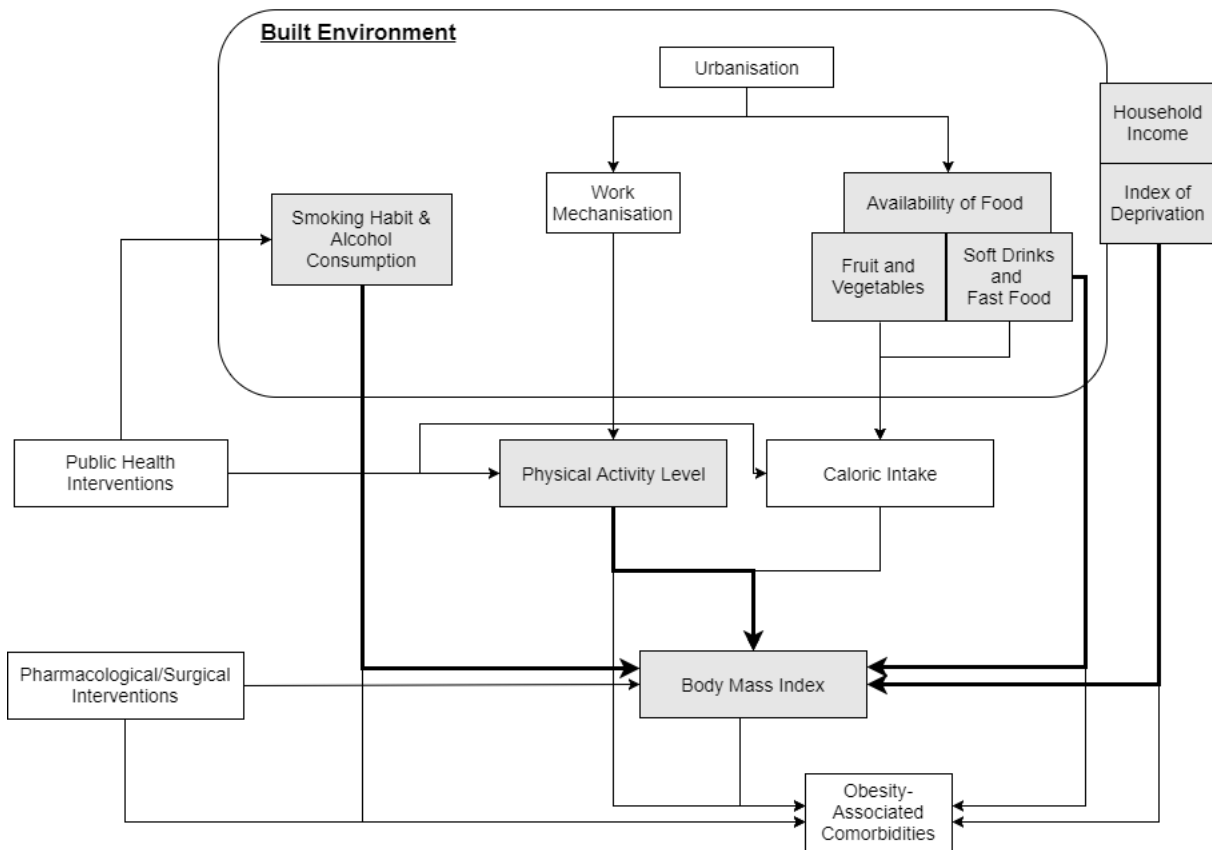


Figure 3.1: Theoretical Framework. Urbanisation is thought to change the built environment, which in turn modifies the behaviours of the people. Work mechanisation and the changes in the availability of certain food may influence physical activity levels and the caloric intake of the population, all of which would impact on the BMI. Also, income and area level deprivation through their relationship with the built environment are thought to have an effect on BMI and obesity-associated comorbidities. Grey boxes indicate available data and bold lines are the main focus of this study.

3.2 Data

I obtained data on age, ethnicity, weight, height, smoking status, drinking problem, physical activity level and deprivation quintile from the confidentialised unit record files (CURFs) of NZHSs. The NZHSs were conducted in 1996/97, 2002/03, 2006/07, 2011/12, 2012/13, 2013/14 and 2014/15. This study excluded the NZHS 1996/97 from the analysis because it did not have the information on the height and weight of individuals. I dropped the data on individuals below 15 years old from the Child Nutrition Survey 2002 (which was intended to supplement the NZHS 2002/03) because it used a different set of questions.

The NZHS is intended to capture the health status of a representative sample of New Zealand residents, which is generalisable to the New Zealand population. It is administered to individuals aged 15 years and above, as well as to children 0 to 14 years, generally through their primary caregiver. The sampling design is multi-stage, stratified, probability-proportional-to-size with meshblocks as the primary sampling units. A meshblock

is the smallest geographically defined area unit for which statistical data is reported by Statistics New Zealand. Statistical weighting is used to ensure the data are representative of the target population, including the sub-groups of population. The content of the health surveys has been preserved so that the data can be compared over time [Ministry of Health, 2015b]. After excluding respondents younger than two years old, the total sample size obtained from the NZHSs was 95,034 respondents.

3.3 Hypotheses

Area level deprivation is a variable used to measure the degree of deprivation a certain area has. In New Zealand, it is calculated using the following information: communication access, income support, income level, employment, educational qualification, home ownership, supports, living space, and means of transportation. The New Zealand Deprivation Quintile is a relative measure of area level deprivation which ranges from one (the least deprived) to five (the most deprived) [Atkinson et al., 2014]. Higher deprivation quintile is associated with a higher overweight/obesity prevalence [Ministry of Health, 2015a, Ministry of Health, 2016]. Thus, I formulated my hypotheses as follows:

- **Hypothesis 1**

Null (H_0): The likelihood of overweight/obesity is the same among residents of areas with different deprivation quintiles.

Alternative (H_1): The likelihood of overweight/obesity is higher among residents of areas with higher deprivation quintiles.

- **Hypothesis 2**

This second set of hypotheses is intended to control for confounding factors.

Null (H_0): The likelihood of overweight/obesity is the same among residents of areas with different deprivation quintiles after adjusting for age, sex, ethnicity, household income, education and health-related behaviours.

Alternative (H_1): The likelihood of overweight/obesity is higher among residents of areas with higher deprivation quintiles after adjusting for age, sex, ethnicity, household income, education and health-related behaviours.

3.4 Variables

3.4.1 Outcome Variables

- Overweight/Obese

The parameter used to measure overweight and obesity is the BMI. BMI is a person's weight in kilograms divided by the square of height in meters. In adults, overweight is defined by a BMI greater than or equal to 25, and obesity is defined by a BMI greater than or equal to 30 [Centers for Disease Control and Prevention, 2013]. The International Obesity Task Force (IOTF) BMI cut-offs were used to define overweight and obesity in participants aged 2–17 years [Ministry of Health, 2015b, Cole et al., 2000, Cole and Lobstein, 2012]. The IOTF BMI cut-offs table can be obtained from the World Obesity Federation (formerly known as IOTF) website [World Obesity Federation,].

3.4.2 Explanatory Variables

- New Zealand Deprivation Quintile

That is an ordinal scale from one (least deprived) to five (most deprived) which measures nine dimensions of area level deprivations.

3.4.3 Covariates

- Physically Active

This is a dichotomous variable which tells whether people aged 18 years or above had 30 minutes of brisk walk or moderate physical exercise (15 minutes if it is a vigorous physical activity) on five or more days in the past seven days.

- Sedentary Lifestyle

This is a dichotomous variable that tells whether people aged 18 years or above had not had at least 30 minutes of exercise in the past seven days.

- Fruit Guideline

This is a dichotomous variable that tells whether people met the dietary fruit guideline, which is two or more servings of fruit per day.

- Vegetable Guideline

This is a dichotomous variable that tells whether people met the dietary vegetable guideline, which is three or more servings of vegetable per day.

- Migration Status

This is a dichotomous variable that tells whether people aged 18 years or more had migrated to New Zealand less than 10–11 years ago at the time of the survey.

- Urban or Rural Area

Urban area is defined as an area that has a population of at least 1000 people. Rural area is defined as an area that has a population of less than 1000 people.

- Educational Qualification

People who had either secondary, tertiary, or no educational qualification. In children, it refers to the educational qualification achieved by the parents, whomever the highest.

- Difficulty Climbing Several Flights of Stairs

People aged 18 years or more who had either a lot of difficulties, a little difficulties or no difficulty climbing several flights of stairs.

- Smoking status

People aged 18 years or above who were either non-smokers, ex-smokers, or current smokers. Current smokers were people who had smoked at least 100 cigarettes over a life-time and were currently smoking at least monthly. Ex-smokers were people who had smoked more than 100 cigarettes over a life time and had stopped for more than a month from the date of interview. Non-smokers were people who were not elsewhere included.

- Drinking Problem

People who had an Alcohol Use Disorders Identification Test (AUDIT) score greater than seven. AUDIT is a screening tool developed by the WHO to assess alcohol consumption, drinking behaviors, and alcohol-related problems. A score greater than seven is considered to indicate harmful alcohol use.

- Soft Drink Consumption

People aged 2–14 years who consumed soft drink in one of the following patterns: none, one, two to three or more than four times per week. Soft drink here includes cola, lemonade or sugar-sweetened energy drinks; but excludes powdered cordial made with water or plain fruit juice. The emphasis is on sugar-sweetened beverages.

- Fast Food Consumption

People aged 2–14 years who consumed fast food in one of the following patterns: none, one, two to three, or more than four times per week. Fast food here includes fish and chips, pizza, burger, fried chicken or other takeaways.

- Household Income

People who belonged to one of the four household income categories: $\leq \$15,000$; $\$15,001$ to $\$40,000$; $\$40,001$ to $\$70,000$; and $\geq \$70,001$.

- Year

That is the period when the New Zealand Health Surveys were conducted.

- Age

It is a numerical scale derived from the date of birth and the date of the interview in years.

- Sex

People who identified as either male or female.

- Ethnicity

People who identified themselves as belonging to Māori only, Pacific only, Asian only (Indian and/or Chinese), European only, 2+ ethnicities (M) (belonging to two or more ethnic group with Māori ancestry), or Other (not elsewhere included).

3.5 Data Analysis

3.5.1 Data Cleaning

Overweight and obesity categories in adults were defined using the WHO criteria. According to this, an individual is defined as overweight if the BMI score is between ≥ 25 and 30 kg/m^2 , and obese if the BMI score is $\geq 30 \text{ kg/m}^2$. 1.2kg is added to the weight variable to remove the clothing adjustment made in the NZHS 2002/03. The existing overweight and obesity category in the NZHS 2002/03 used higher cut-offs for Māori and Pacific people. In the later surveys, the WHO overweight/obesity cut-offs were used for everyone regardless of their ethnicity. New code was written to create overweight and obesity category according to the WHO criteria for NZHS 2002/03 so that the data would be comparable to the rest of the health surveys.

The cut-offs used for child overweight and obesity in the NZHS 2006/07, 2011/12, and 2012/13 were based on the old IOTF cut-offs. Even though the old and new IOTF cut-offs do not differ significantly in defining overweight and obesity [Cole and Lobstein, 2012], this study decided to use the new IOTF cut-offs based on the children's age in years across the data sets to be consistent. The new IOTF cut-offs based on the children's age in years were also applied to the NZHS 2013/14 and 2014/15, because they used the cut-offs based on the children's age in months.

Children aged 0–14 years are recorded in the child NZHS, while 15+ years old are recorded in the adult NZHS. I combined and then split the data into the Child data (2–17 years old) and the Adult data (18+ years old).

I did this so that the analysis on adults would be easier as the overweight/obesity criteria does not vary by age starting from 18 years old.

Imputations on missing income data were only done in the NZHS 2002/03, but not in any of the other survey periods. Therefore, I decided not to use any imputation in my analysis. Imputed income data in the NZHS 2002/03 were removed and the missing income data from the rest of the data sets were excluded list-wise from the analysis. Although there was no household income data and parent's educational level in the NZHS 2011/12 and 2012/13 child data, this information could be obtained through matching the household ID (HHID) from the adult data of the same survey period. After matching the HHID, there were two different educational level data for the same child. This discrepancy was not addressed in any of the NZHS methodology reports published by the Ministry of Health and may have happened because there were two parents who had different qualifications. I used the highest parental education level in this case.

The NZHSs used ethnic prioritisation in their ethnicity variable, which means if someone identified themselves as having multiple ethnicities, the ethnicity variable would have been forced into one ethnicity with a higher priority. The order of priority was Māori, Pacific, Asian, and European/other ethnicity. For example, people who identified as belonging to Māori and Asian descents would be treated as being in the Māori only group. Even though ethnic prioritisation can simplify the analysis, it is not ethnically neutral and eliminates the information that can be gained from younger people who more frequently belong to multiple ethnic groups [Didham and Callister, 2012]. Codes for a new ethnic groupings were written, and the new groupings were broken down into Māori only, Pacific only, Asian only, European only, 2+ ethnicities (M), and "Other" group. Other Asians (besides Indian and/or Chinese) were included in the "Other" category because there was no option for the respondents to choose other Asian ethnicities in the NZHSs.

The R codes used to clean the NZHS data are presented in Listing A.1.

3.5.2 Descriptive and Bivariate Analyses

I conducted descriptive statistical analysis of single variables after applying the sample weights. The nominal or ordinal variables were reported as count and percentages while the numeric variables were reported as five numbers summary (mean, minimum, maximum, first inter-quartile and third inter-quartile) in a table. Unweighted bivariate analyses were conducted on all unique combinations between explanatory variables and the outcome variable. Chi-squared test was used when both the explanatory and the outcome variable were categorical. ANOVA test was used when the explanatory variable was numeric and the outcome variable was categorical. All statistical results from the bivariate analyses along with their effect sizes were reported. I also conducted bivariate analyses on each survey period separately and the R codes are presented in Listing A.3 and A.4. In addition, I plotted the prevalence of overweight/obesity over time, and the age-standardised obesity prevalence by ethnicity

over time.

3.5.3 Regression Modelling

The analysis in this study used the sample weights so that the results would be representative to the whole population. The NZHS 2006/07 did not have unique identifiers for its cluster variable, thus unique clusters were created for the NZHS 2006/07 as to be consistent with the remaining health surveys. The complex survey design object was created using the 'survey' package in R [Lumley, 2016, Lumley, 2004]. The R codes used to do the ordinal regression analyses are presented in Listing A.2.

A forward stepwise proportional odds ordinal regression using complex survey design was conducted to find the final model. My interest was in the effect of area level deprivation on overweight/obesity outcome. In the first step, deprivation quintile and linear contrast were added into the model. Then, I added one covariate at a time and dropped the covariate if it did not reach a p-value of at least 0.1. Proportional odds assumption was assessed by running a binomial general linear model with contrasts on different threshold values. I followed the instruction to assess proportional odds assumption by 'coreysparks' [coreysparks, 2015] and the R codes were also presented in Listing A.2.

The proportional odds of cumulative logit model is given in the following equation [Agresti, 2013]:

$$\text{logit}[P(Y \geq j|x_1, \dots, x_i)] = \alpha_j + \beta_1^T x_1 + \dots + \beta_i^T x_i, \quad (j = 1, \dots, J - 1; i = 1, \dots, n) \quad (3.1)$$

Where:

Y = outcome variable

j = threshold point which divide Y into two categories

x = explanatory variable

α_j = intercept coefficient for j threshold

β_i^T = beta coefficient for each explanatory variable x , assumed to be the same for each logit

$P[Y \geq j|x_1, \dots, x_i]$ = probability of Y outcome above j threshold given explanatory variable x_1, \dots, x_i

4 Results

4.1 Descriptive Analysis of The NZHS

The descriptive summary of the Adult data is presented in table 4.1 after applying the sample weights. The prevalence of obesity had not changed statistically significantly since 2011/12. The prevalence in 2014/15 (32%) was higher than in 2002/03 (26%) ($p < .001$) and in 2006/07 (28%) ($p < .001$) (Figure 4.1). In regards to ethnicity, the proportion of European group decreased while the Other group increased over time. The proportion of respondents who lived in urban areas was higher in 2014/15 (86.8%) than in 2002/03 (80.1%) ($p < 0.01$) and 2006/07 (86%) ($p < .001$). There is no data on the urban/rural information in the 2011/12 and 2013/14 NZHS, thus no conclusion could be made on those years. In regards to health-related behaviours, the proportion of people who met the fruit guideline in 2014/15 (55%) was lower than the figure in 2006/07 (60%) ($p < .001$), and it was not different than in 2002/03 (54%) ($p = .2$). The proportion of people who met the vegetable guideline decreased from 69% in 2011/12 to 66% in 2014/15 ($p < .001$). There was no statistically significant difference in the proportion of people who were physically active across the survey periods except in 2011/12 where it peaked at 56%. The proportion of people who lived a sedentary lifestyle in 2014/15 (14%) was higher compared with the figure in 2011/12 (11%) ($p < .001$), but it was lower than in 2002/03 (16%) ($p < .01$). The BMI in the Adult data had a skewness of 1.26 and a kurtosis of 5.62 ($p < .001$), which means the BMI data was skewed to the right and leptokurtic. Therefore, I categorised the BMI variable in the Adult data into tertiles.

A summary of descriptive statistics for the Child data is presented in Table 4.2. The proportion of obese children in 2012/13 (13%) was higher than in 2006/07 (10%) ($p = .014$). However, compared with the proportion of obese children in 2014/15 (13%), there was no difference since 2011/12 ($p < .001$) (Figure 4.1). Similar to the Adult data, the proportion of the European group also decreased while the proportion of the Other ethnic group increased over time. The proportion of children who lived in urban areas was lower in 2014/15 (80%) compared with the figure in 2006/07 (85%) ($p = .01$). The proportion of parents who had tertiary education was higher in 2014/15 (44%) compared with the figure in 2012/13 (39%) ($p < .01$). There was no significant statistical difference in the proportion of children who had soft drink or fast food more than once per week across time. Missing data in both Adult and Child data are presented in Table 4.3.

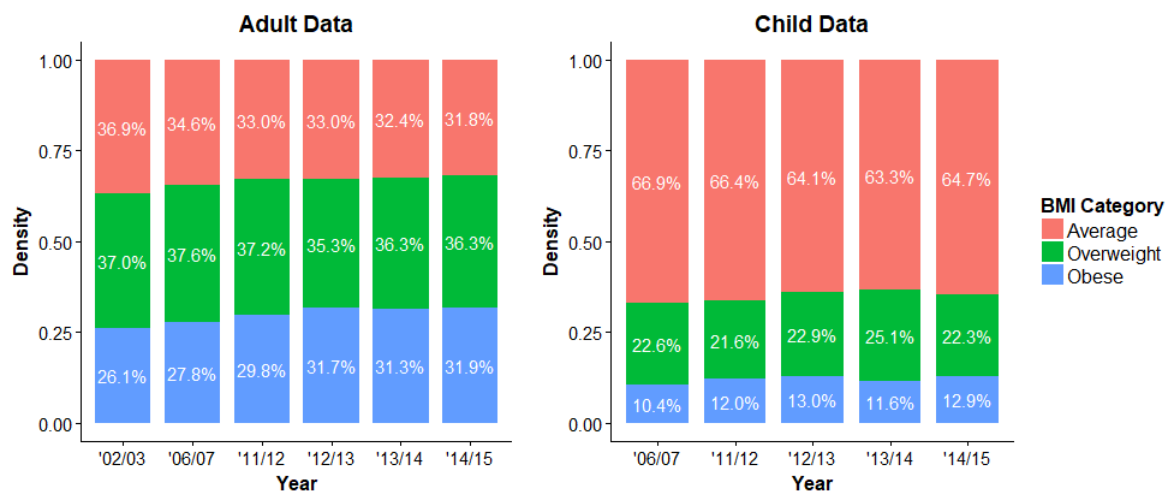


Figure 4.1: BMI Categories in The Adult and Child Data

Table 4.1: Descriptive Statistics of The Adult Data (2002/03–2014/15 NZHS)

Variables		New Zealand Health Survey Period					
		2002/03	2006/08	2011/12	2012/13	2013/14	2014/15
Sample Size		10812	11010	10234	11444	12046	12185
Age (in Years)	Minimum	18	18	18	18	18	18
	1st Quartile	31	33	32	32	32	32
	Median	43	44	46	46	46	46
	3rd Quartile	56	58	59	60	61	61
	Max	97	95*	90*	90*	90*	90*
BMI (kg/m²)	Minimum	18.53	18.51	18.50	18.50	18.50	18.50
	1st Quartile	23.59	23.81	24.06	23.99	24.12	24.16
	Median	26.54	26.86	26.98	27.18	27.31	27.32
	3rd Quartile	30.20	30.53	30.91	31.38	31.21	31.27
	Max	64.21	80.56	77.01	68.20	70.86	78.02
		% (SD)					
BMI Category	Average	36.88 (0.70)	34.60 (0.64)	32.98 (0.74)	32.96 (0.69)	32.40 (0.66)	31.80 (0.69)
	Overweight	37.02 (0.65)	37.58 (0.59)	37.21 (0.69)	35.32 (0.66)	36.31 (0.62)	36.34 (0.62)
	Obese	26.09 (0.64)	27.82 (0.58)	29.80 (0.69)	31.72 (0.65)	31.30 (0.68)	31.86 (0.68)
Sex	Male	49.99 (0.65)	49.72 (0.7)	49.84 (0.64)	50.87 (0.69)	49.74 (0.63)	49.34 (0.62)
Ethnic Group	European Only	78.72 (0.87)	73.93 (0.67)	60.30 (0.90)	62.43 (0.85)	62.41 (0.85)	60.86 (0.85)
	2+ Ethnicities (M)	3.93 (0.27)	5.30 (0.22)	5.38 (0.28)	5.02 (0.26)	5.17 (0.25)	5.54 (0.30)
	Asian Only	3.82 (0.33)	5.83 (0.35)	5.46 (0.49)	5.80 (0.42)	6.27 (0.46)	6.95 (0.49)
	Maori Only	6.31 (0.38)	5.58 (0.26)	6.24 (0.34)	7.27 (0.36)	6.97 (0.36)	6.69 (0.33)
	Pacific Only	3.96 (0.45)	4.12 (0.34)	3.21 (0.30)	3.62 (0.32)	3.37 (0.30)	3.63 (0.30)
	Other	3.26 (0.27)	5.24 (0.28)	19.41 (0.66)	15.86 (0.55)	15.79 (0.54)	16.33 (0.53)
Deprivation Quintile	1	20.45 (1.88)	18.4 (1.31)	21.73 (1.34)	20.72 (1.84)	20.78 (1.76)	20.24 (1.83)
	3	20.91 (1.64)	20.88 (1.5)	20.56 (1.2)	20.63 (1.68)	20.32 (1.6)	20.75 (1.5)
	5	18.45 (1.19)	19.39 (1.5)	17.62 (1.08)	18.04 (1.3)	18.24 (1.26)	18.29 (1.22)
Urban/Rural Area	Urban	80.13 (1.45)	86.01 (1.25)	-	-	-	86.75 (0.96)
Household Income	<= \$15k	10.1 (0.49)	5.87 (0.28)	3.55 (0.3)	4.16 (0.36)	3.78 (0.28)	3.69 (0.26)
	\$15–40k	32.75 (0.85)	25.15 (0.56)	21.19 (0.7)	23.07 (0.69)	21.04 (0.67)	21.88 (0.67)
	\$40–70k	27.03 (0.69)	25.52 (0.56)	25.11 (0.7)	24.97 (0.67)	24.19 (0.68)	24.27 (0.67)
	>\$70k	30.12 (0.91)	43.46 (0.78)	50.15 (1.03)	47.81 (0.94)	50.99 (0.92)	50.16 (0.97)
Educational Qualification	None	20.73 (0.65)	16.88 (0.47)	14.81 (0.52)	15.64 (0.51)	14.06 (0.48)	13.89 (0.46)
	Secondary	28.36 (0.64)	26.41 (0.58)	31.68 (0.68)	32.45 (0.66)	34.45 (0.65)	33.49 (0.67)
	Tertiary	50.91 (0.8)	56.71 (0.67)	53.51 (0.83)	51.91 (0.76)	51.49 (0.75)	52.62 (0.77)
Meeting Fruit Guideline	Yes	54.1 (0.75)	59.6 (0.63)	58.44 (0.72)	58.09 (0.69)	56.42 (0.67)	55.37 (0.67)
Meeting Vegetable Guideline	Yes	69.14 (0.83)	65.04 (0.68)	69.23 (0.79)	67.44 (0.8)	64.91 (0.76)	65.52 (0.74)
Migration Status	Migrant	4.46 (0.31)	8.67 (0.39)	11.13 (0.56)	10.26 (0.5)	10.63 (0.53)	11.01 (0.53)
Difficulty Climbing Several Flights of Stairs	A lot	8.69 (0.4)	9.64 (0.33)	5.66 (0.3)	6.46 (0.28)	7.04 (0.32)	7.02 (0.32)
	A little	13.5 (0.48)	16.82 (0.45)	10.4 (0.4)	10.95 (0.4)	10.06 (0.38)	11.19 (0.43)
	None	77.81 (0.48)	73.53 (0.45)	83.94 (0.4)	82.59 (0.4)	82.89 (0.38)	81.79 (0.43)
Smoking Status	Non Smoker	-	55.5 (0.7)	54.86 (0.73)	55.13 (0.69)	55.11 (0.67)	55.74 (0.66)
	Ex Smoker	-	24.14 (0.54)	26.27 (0.55)	26.8 (0.57)	26.97 (0.56)	27.23 (0.58)
	Current Smoker	-	20.36 (0.54)	18.87 (0.58)	18.07 (0.54)	17.91 (0.53)	17.03 (0.52)
Drinking Problem	AUDIT >7	17.5 (0.63)	18.13 (0.54)	15.74 (0.58)	16.06 (0.55)	16.81 (0.57)	18.53 (0.61)
Physically Active	Yes	53.12 (0.81)	52.85 (0.68)	56.18 (0.96)	52.77 (1.01)	53.03 (1)	51.55 (0.82)
Sedentary Lifestyle	Yes	16.21 (0.61)	13.22 (0.4)	10.86 (0.5)	12.98 (0.53)	13.36 (0.52)	13.75 (0.47)

* Maximum age is capped at this number

Table 4.2: Descriptive Statistics of The Child Data (2006/07–2014/15 NZHS)

Variables		New Zealand Health Survey Period				
		2006/08	2011/12	2012/13	2013/14	2014/15
Sample Size		4415	3310	3415	3788	4020
Age (in Years)	Minimum	2	2	2	2	2
	1st Quartile	6	6	6	6	6
	Median	10	10	10	10	10
	3rd Quartile	14	14	14	14	14
	Maximum	17	17	17	17	17
		% (SD)				
Sex	Male	51.61 (1.01)	52.49 (1.26)	50.68 (1.16)	51.95 (1.17)	50.32 (1.11)
BMI Category	Average	66.94 (1)	66.39 (1.27)	64.1 (1.19)	63.35 (1.08)	64.73 (1.18)
	Overweight	22.65 (0.86)	21.6 (0.96)	22.95 (1.03)	25.07 (0.92)	22.33 (0.94)
	Obese	10.41 (0.61)	12.01 (0.79)	12.96 (0.84)	11.58 (0.7)	12.94 (0.72)
Ethnic Groups	European Only	60.33 (1.14)	48.08 (1.57)	51.02 (1.53)	46.81 (1.42)	48.93 (1.46)
	2+ Ethnicities (M)	13.29 (0.58)	12.35 (0.73)	13.55 (0.78)	15.05 (0.78)	14.49 (0.76)
	Asian Only	4.01 (0.37)	4.53 (0.63)	4.5 (0.57)	5.11 (0.6)	4.48 (0.51)
	Maori Only	8.77 (0.5)	10.29 (0.84)	11.03 (0.77)	10.53 (0.71)	11.49 (0.75)
	Pacific Only	7.19 (0.68)	7.57 (0.92)	6.81 (0.76)	5.9 (0.65)	5.65 (0.57)
	Other	6.42 (0.68)	17.18 (0.92)	13.07 (0.76)	16.59 (0.65)	14.94 (0.57)
Deprivation Quintiles	1	21.87 (1.51)	22.11 (2.16)	20.59 (1.97)	20.95 (2.13)	19.83 (2.14)
	3	19.38 (1.22)	16.7 (1.55)	18.74 (1.63)	17.13 (1.52)	17.52 (1.55)
	5	20.49 (1.27)	23.41 (1.82)	23.29 (1.62)	23.68 (1.61)	23.69 (1.58)
Urban/Rural	Urban	85.06 (1.15)	-	-	-	80.06 (1.67)
Household Income	<= \$15k	4.68 (0.42)	3.78 (0.63)	6.2 (1.33)	4.06 (0.56)	3.21 (0.45)
	\$15–40k	23.15 (0.94)	16.92 (1.32)	17.85 (1.18)	17.83 (1.03)	18.04 (1.03)
	\$40–70k	29.91 (1.05)	23.71 (1.43)	25.64 (1.28)	26.35 (1.32)	23.06 (1.07)
	>\$70k	42.26 (1.23)	55.59 (1.86)	50.31 (1.76)	51.76 (1.57)	55.69 (1.48)
Parents Qualification*	None	19.34 (0.86)	19.54 (1.22)	21.3 (1.13)	14.43 (0.92)	15.64 (0.88)
	Secondary	35.74 (1.07)	40.12 (1.41)	40 (1.27)	42.31 (1.27)	40.71 (1.19)
	Tertiary	44.93 (1.09)	40.35 (1.46)	38.7 (1.37)	43.26 (1.2)	43.65 (1.24)
Meeting Dietary Guideline	Fruit	61.44 (2.79)**	68.48 (1.2)	71.01 (1.16)	69.38 (1.09)	70.87 (1.04)
	Vegetable	53.01 (2.76)**	60.05 (1.33)	58.41 (1.28)	57.44 (1.27)	58.04 (1.19)
Soft Drink Consumption (per week)*	0	36.47 (1.08)	36.78 (1.29)	36.9 (1.16)	40.47 (1.25)	39.74 (1.25)
	1	28.03 (0.94)	28.48 (1.12)	28.39 (1.12)	27.68 (1.07)	27.65 (1.03)
	2–3	23.62 (0.89)	21.44 (1.07)	23.01 (1.14)	21.36 (0.93)	20.37 (1.02)
	4+	11.89 (0.68)	13.3 (0.9)	11.71 (0.69)	10.5 (0.68)	12.25 (0.75)
Fast Food Consumption (per week)*	0	28.92 (0.98)	31.49 (1.12)	33.11 (1.12)	32.36 (1.25)	30.45 (1.07)
	1	50.75 (1.03)	48.22 (1.23)	47.59 (1.22)	47.85 (1.26)	49.19 (1.13)
	2–3	16.95 (0.78)	17.82 (0.95)	17.08 (0.85)	17.79 (0.88)	18.1 (0.89)
	4+	3.38 (0.4)	2.47 (0.4)	2.23 (0.31)	2 (0.31)	2.27 (0.29)

* Below 15 years old

** Data on children younger than 15 years old is not available

Table 4.3: Missing Data Table

Missing Data*	2002/03	2006/07	2011/12	2012/13	2013/14	2014/15
N (%)						
Adult BMI	1189 (9.78)	788 (6.68)	1864 (15.41)	1038 (8.32)	696 (5.46)	713 (5.53)
Child BMI	-	177 (3.86)	668 (16.77)	543 (13.72)	372 (8.94)	301 (6.97)
Adult Household Income	2807 (23.09)	1393 (11.81)	3920 (32.40)	3651 (29.25)	3639 (28.56)	2220 (17.21)
Child Household Income	-	614 (13.37)	1612 (40.47)	1445 (36.51)	1379 (33.15)	958 (22.17)
Parent's Education	-	-	194 (4.87)	246 (6.22)	195 (4.69)	167 (3.87)

* Only showing variables with 3% or more missing data

4.2 Bivariate Analyses

I did not use sample weights in these analyses. Figure 4.2 and Figure 4.3 show the bivariate charts on all possible combinations between the covariates and the BMI outcomes, which are BMI tertiles in adults and BMI categories in children. The bivariate tables are presented in Appendix B. Summary of the bivariate statistical tests are presented in Appendix B.3. Even though most explanatory variables were statistically significantly associated with the outcome variables, only several variables had a small or medium effect size according to the Cohen's guide [Cohen, 1992]. Age ($\eta^2 = 0.01$), sex ($\varphi_c = 0.11$) and difficulty climbing stairs ($\varphi_c = 0.11$) had a small size effect in the Adult data, while deprivation quintile and education had a small effect size in both the Adult ($\varphi_c = 0.12$; $\varphi_c = 0.09$) and Child data ($\varphi_c = 0.11$; $\varphi_c = 0.07$). Ethnicity variable was the only variable that had a medium effect size in the Adult ($\varphi_c = 0.22$) and Child data ($\varphi_c = 0.15$). Separate bivariate analyses on each survey period are presented as R codes in Listing A.3 and A.4.

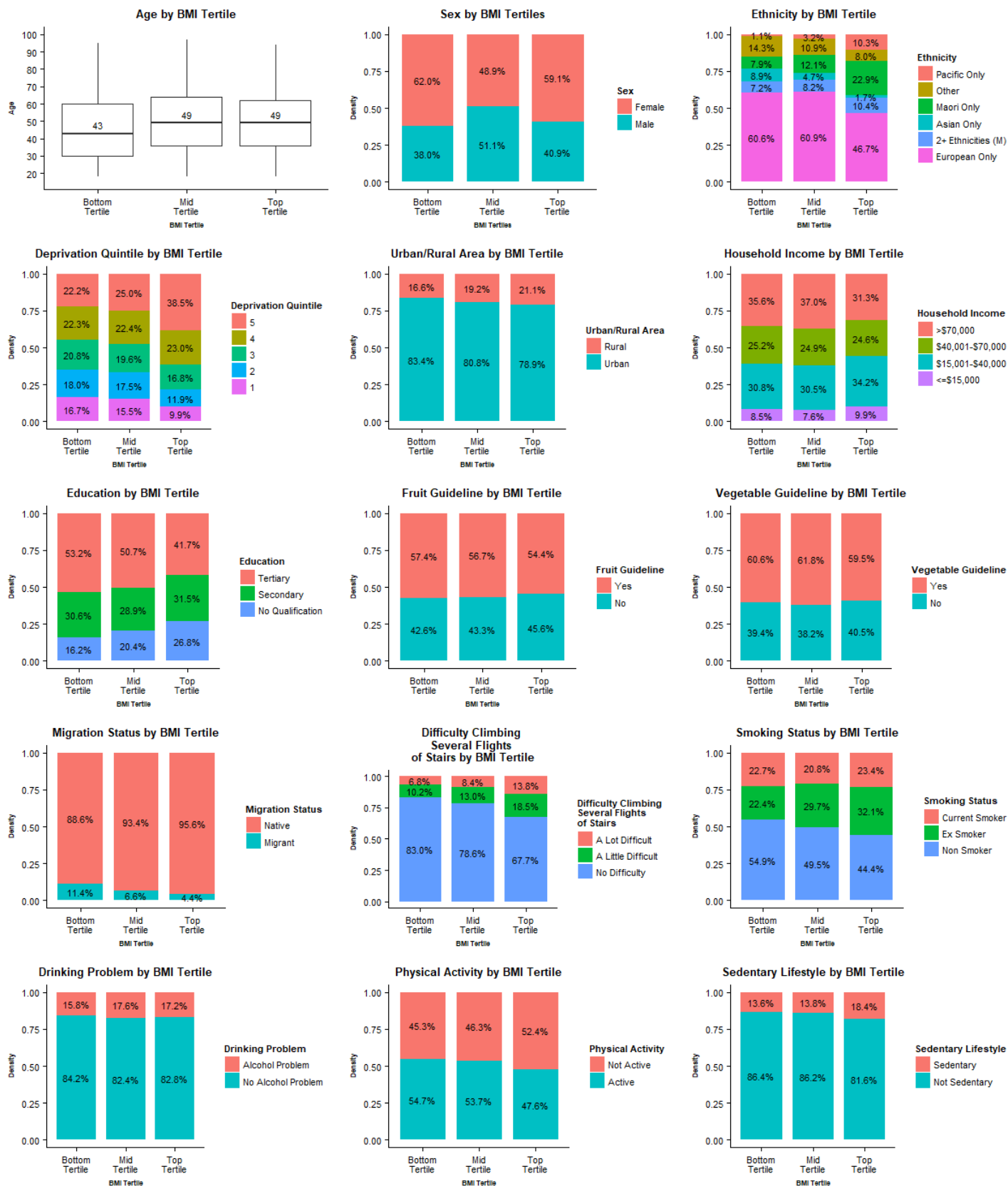


Figure 4.2: Bivariate Charts for The Adult Data

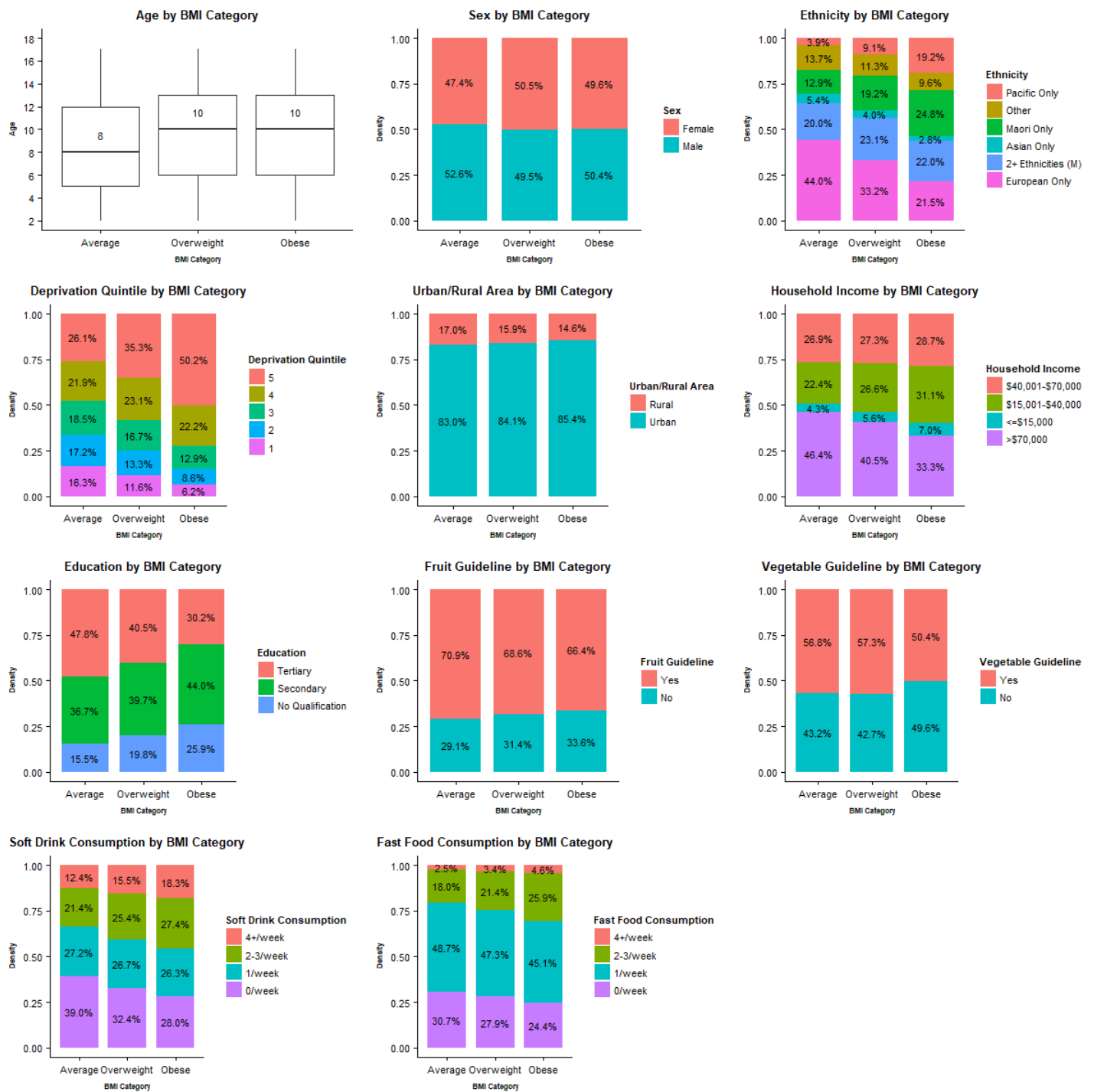


Figure 4.3: Bivariate Charts for The Child Data

4.3 Obesity Prevalence across Ethnic Groups

Figure 4.4 shows ethnic-specific obesity prevalence over time in adults and children. In adults, almost three out of every four Pacific adults were obese in 2014/15 and this was more than twice the proportion of obese European adults (31%). Māori had the second highest obesity prevalence (57%) in the same year and 43% of people who identified themselves as having multiple ethnicities with Māori ancestry were obese. If overweight status was included, more than 90% of Pacific and more than 80% of Māori would be considered overweight/obese using the universal BMI cut-offs over the survey periods (Figure 4.5). There was an overall increasing trend of adult obesity prevalence over time.

Similar to adults, Pacific and Māori children had a higher obesity prevalence compared to the other ethnic groups. Almost half of Pacific and one in every four Māori were obese in 2014/15. On the other hand, only 7% of Asians and one in eleven Europeans were obese in the same year. The obesity prevalence among Pacific children had grown from 34% in 2006/07 to 47% in 2014/15, whereas the prevalence had decreased among Asians and remained stable in other ethnic groups in the same period. Even though the obesity prevalence among Asians decreased from 9% in 2006/07 to 7% in 2014/15, the proportion of overweight Asian children increased; this equates to an overall increase in overweight/obesity prevalence among Asians in the same period. Furthermore, if overweight status was included, more than six out of ten Pacific and around half of Māori children would be considered overweight/obese throughout the survey periods (Figure 4.5).

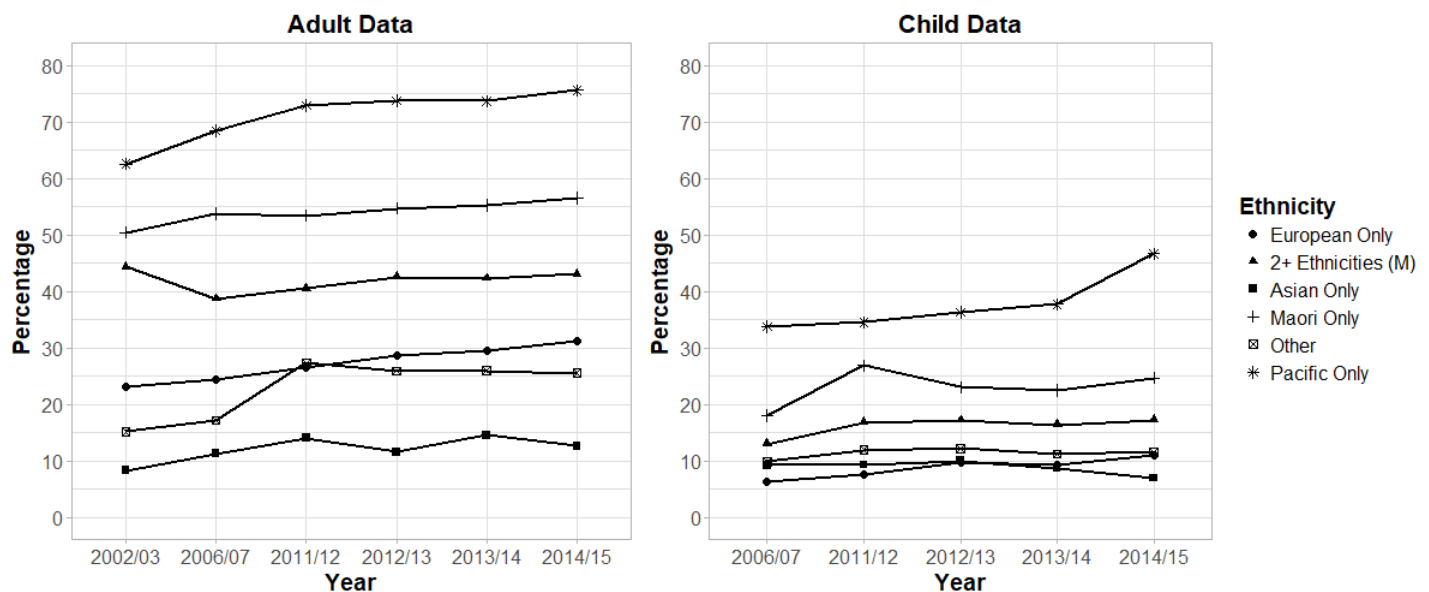


Figure 4.4: Age-Standardised Obesity Prevalence by Ethnicity from 2002/03 to 2014/15

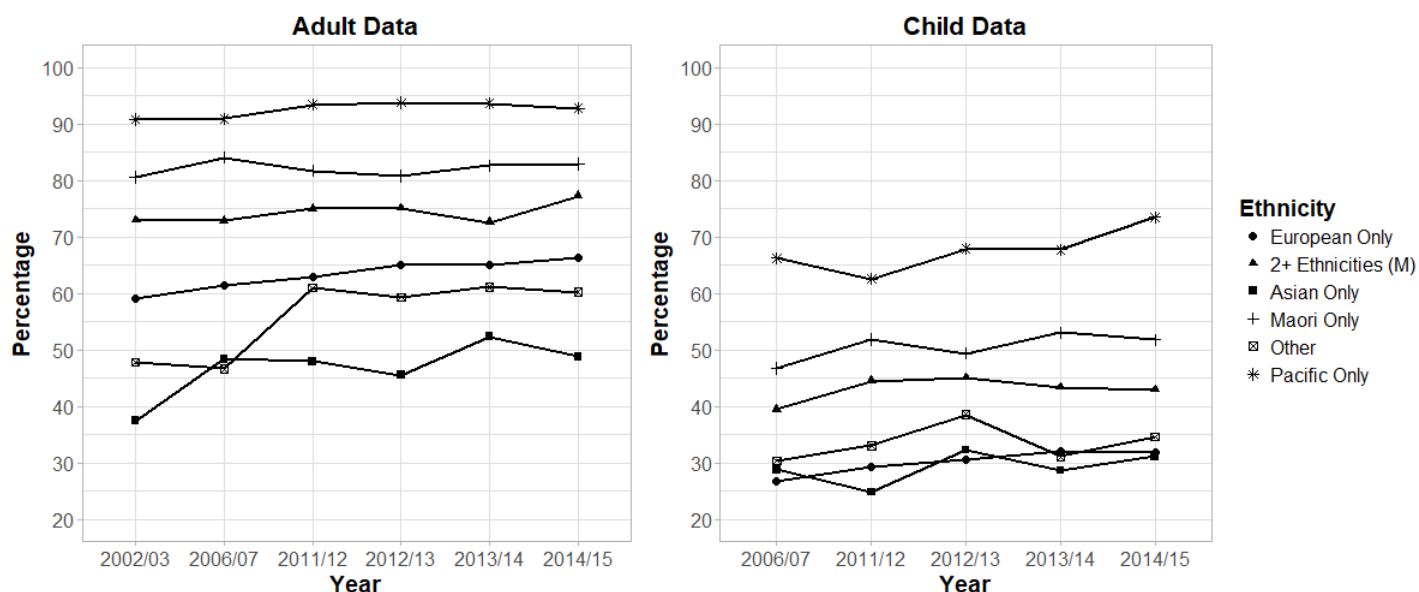


Figure 4.5: Age-Standardised Overweight/Obesity Prevalence by Ethnicity from 2002/03 to 2014/15

4.4 Regressions Result

I conducted a forward stepwise proportional odds regression with complex survey design in the Adult data to determine the final model. Before controlling for other covariates, people who lived in quintile five had 90% more risk of being in the higher BMI tertiles than those who lived in quintile one ($OR = 1.89$, $95\%CI = 1.75-2.05$). Linear contrast was then added to the model and was significant ($OR = 1.21$, $95\%CI = 1.15-1.28$); indicating that there was a secular change in BMI tertiles over time. Meeting vegetable guideline ($OR = 1.03$, $95\%CI = 0.99-1.07$) and fruit guideline ($OR = 0.97$, $95\%CI = 0.93-1.005$) did not exert a significant effect on BMI tertiles after adjusting for deprivation quintiles and linear contrast. On the same step, drinking problem was not a good predictor of BMI tertiles ($OR = 0.99$, $95\%CI = 0.94-1.04$). Sedentary lifestyle also lost its significance after adjusting for deprivation, linear contrast, ethnicity, age and difficulty climbing several flights of stairs ($OR = 1.06$, $95\%CI = 0.99-1.12$). There was no difference in BMI tertiles between people who lived in urban and rural areas after controlling for smoking status, sex, difficulty climbing stairs, age, ethnicity, and linear contrast ($OR = 1.07$, $95\%CI = 0.98-1.17$).

The summary of the final regression results are presented in Table 4.4. Having a higher educational qualification was protective against being in the higher BMI tertiles. Similarly, current smokers were less likely to be in the higher BMI tertiles, which indicates the 'nicotine effect'. All other variables in the model increased the likelihood of being in the higher BMI tertiles. The odds of being in the higher BMI tertiles increased as the level of area deprivation increased. Higher household income was associated with a higher risk of obesity, this finding is unusual considering a positive relationship between area level deprivation and obesity. Ethnicity remained a strong predictor of obesity even after adjusting for other variables.

I conducted another forward stepwise proportional odds regression with complex survey design to build a model for the Child data. The effect of meeting fruit guidelines (OR = 0.93, 95%CI = 0.84-1.03) and vegetable guidelines (OR = 0.99, 95%CI = 0.91–1.09) were not significant after adjusting for deprivation level and linear contrast. There was no evidence that urban/rural area had an influence on BMI categories after controlling for deprivation level and linear contrast (OR = 0.96, 95%CI = 0.8–1.14). Unlike the finding in adults, the BMI outcome of children who lived in the poorest household income (\leq \$15,000) was not significantly different than those in the richest household income ($>$ \$70,000) after adjusting for ethnicity and deprivation level (OR = 1.23, 95%CI = 0.96-1.59). Fast food consumption was not a significant predictor of BMI outcome (4+/week versus $<$ 1/week; OR = 1.08, 95%CI = 0.83–1.39) after adjusting for soft drink consumption, education, age, ethnicity, and deprivation; this may suggest that there was a multicollinearity in the relationship between soft drink and fast food consumption.

The final regression results are presented in Table 4.5. Higher parents' educational qualification protected the children against overweight/obesity. Ethnicity remained a strong predictors of BMI outcomes after adjusting for demographic variables and soft drink consumption. The effect of ethnicity, however, was not as strong as in adults. Higher deprivation level equated to higher likelihood of being in the higher BMI categories in children, and the effect was stronger compared to the finding in adults.

Table 4.4: The Regression Summary of The Adult Data

	Odds Ratio	2.50%	97.50%
Deprivation Quintile (ref. group: Quintile 1)			
Quintile 2	1.02	0.94	1.11
Quintile 3	1.10	1.01	1.20
Quintile 4	1.29	1.19	1.40
Quintile 5	1.46	1.33	1.60
Linear Contrast	1.19	1.11	1.28
Ethnicity (ref. group: European only)			
2+ ethnicities (M)	1.72	1.57	1.88
Asian only	0.53	0.48	0.60
Māori only	3.01	2.75	3.29
Pacific only	8.41	7.22	9.79
Other	0.91	0.84	0.98
Age	1.01	1.01	1.01
Difficulty Climbing Several Flights of Stairs (ref. group: No difficulty)			
A little difficult	1.81	1.68	1.96
A lot difficult	1.80	1.64	1.98
Sex			
Female	0.76	0.73	0.80
Smoking Status (ref. group: Non smoker)			
Ex smoker	1.36	1.29	1.44
Current smoker	0.93	0.86	1.00
Household Income (ref. group: ≤\$70,000)			
\$15,001–\$40,000	1.19	1.07	1.32
\$40,001–\$70,000	1.41	1.26	1.57
>\$70,000	1.59	1.42	1.77
Educational Qualification (ref. group: No qualification)			
Secondary	0.89	0.83	0.96
Tertiary	0.80	0.74	0.86
Physical Activity			
Not physically active	1.15	1.09	1.20
Migration Status			
Native	1.22	1.10	1.34

Table 4.5: The Regression Summary of The Child Data

	Odds Ratio	2.50%	97.50%
Deprivation Quintile (ref. group: Quintile 1)			
Quintile 2	1.13	0.94	1.35
Quintile 3	1.41	1.20	1.67
Quintile 4	1.50	1.28	1.76
Quintile 5	1.76	1.50	2.06
Linear Contrast	1.16	1.05	1.28
Ethnicity (ref. group: European only)			
2+ ethnicities (M)	1.49	1.33	1.67
Asian only	1.04	0.84	1.28
Māori only	2.03	1.77	2.32
Pacific only	3.93	3.31	4.68
Other	1.21	1.05	1.38
Age	1.04	1.03	1.06
Parent's Qualification (ref. group: No qualification)			
Secondary	0.73	0.64	0.83
Tertiary	0.62	0.55	0.71
Soft Drink Consumption (ref. group: <1/week)			
1/week	1.15	1.03	1.29
2-3/week	1.24	1.10	1.39
4+/week	1.213	1.06	1.39
Sex			
Female	1.11	1.02	1.21

4.5 Proportional Odds Assumption

The proportional odds assumption in the Adult data did not hold for sex, where male adults were more likely to belong in the mid or top tertile while female adults were more likely to belong in the top tertile. This relationship can also be seen in the sex by BMI tertiles chart (first row and second column of Figure 4.2), this showed that there were more female in the top tertile and there were more male in the mid tertile. Therefore, interpretation of the sex variable of the proportional odds logistic regression should be done carefully. Furthermore, adults who had difficulties climbing several flights of stairs were more likely to be in the top tertile compared with being in the mid or top tertile. In Pacific adults, the likelihood to be in the mid or top tertile was higher compared with being in the top tertile. In children, the proportional odds assumption did not hold for quintile three to five, Māori and Pacific ethnic group; on all of those variables, the likelihoods of being in the obese category were higher compared with the likelihoods of being in the overweight/obese category. The details of this analyses are presented in Listing A.2 and the binomial regression tables for assessing the proportional odds assumption are presented in Appendix C.

5 Discussion

5.1 Summary of the findings

This study found that the prevalence of obesity in adults and children had not changed significantly since 2011/12. Despite that, I found that there was evidence of linear temporal change over time from the regression analysis. The regression analysis showed that ethnicity was a strong predictor of BMI outcomes. This further asserts the finding from the descriptive statistics showing that Pacific and Māori had the highest obesity prevalence among other ethnic groups. The difference in BMI outcomes between ethnicities was even more obvious in adults, suggesting that the BMI across ethnic groups will keep diverging as people age. The main objective of this study is to assess whether socio-economic deprivation, as measured by area level deprivation, influences the BMI outcome. This study found that the odds of being in the higher BMI groups increased as the level of area deprivation where people reside increased, and the association was stronger in children.

In the next section, I will address the association between the area level deprivation and obesity followed by a discussion regarding an accidental finding from this study about household income and obesity. I will then discuss BMI differences across ethnic groups and the problem of using the universal BMI cut-offs. I will also talk about the findings related to diets and physical activity. Lastly, I will discuss the smoking habit, healthy immigrant effect and living in urban/rural areas in relation to obesity.

5.2 Area-Level Deprivation

This study found that the deprivation quintile influenced one's body mass index after adjusting for household income, age, education level, and ethnicity. This finding is consistent with other studies in the US [Nau et al., 2015], Canada [Matheson et al., 2008], UK [Cummins and Macintyre, 2006] and Sweden [Stafford et al., 2010]. Adults and children who lived in the most deprived communities had a higher trajectory of BMI growth over time [Stafford et al., 2010, Nau et al., 2015]. Moreover, studies have found that people who lived in more deprived communities had poorer access to fruit and vegetables [Eagle et al., 2012, Cummins and Macintyre, 2006, Nau et al., 2015], which would lead to a poorer diet quality [Craig et al., 2016]. Furthermore, children in deprived neighbourhoods had lower opportunities for outdoor play due to inadequate infrastructure as well as safety concerns [Eagle et al., 2012, Nau et al., 2015]. On top of that, structural and social barriers to health care services in more deprived communities put people at a higher risk of being overweight/obese [Li et al., 2014]. This study highlights that even

in a country with a strong welfare system, people who lived in more deprived areas had higher odds of developing overweight/obesity. It does not seem that poor people become more obese because they are being "lazy", instead a lot of social and infrastructure barriers put them in a more disadvantaged position with less resources to combat obesity.

5.3 Household Income Level

The association between socio-economic status and obesity seems to depend on the income level of the country. Socio-economic status was inversely related to obesity rates among high-income countries, but the association was reversed in low-income countries [Wu et al., 2015, Fruhstorfer et al., 2016]. In high-income countries, food insecurity is not prevalent anymore even among low-income families due to a strong welfare system that helps them with foods and daily needs. Nonetheless, they are more susceptible to obesity compared with their well-off counterparts due to a lack of resources to choose a healthier diet and lifestyle, which tend to be more expensive. Conversely, in low-income countries, the association was reversed. High-income families were more likely to be overweight/obese than the low-income families. This association occurred because food availability is still a challenge in low-income families and overweight may be perceived as a sign of wealth [Dinsa et al., 2012, Wu et al., 2015, Fruhstorfer et al., 2016].

In New Zealand, contrary to what would have been expected from a high-income country, this study found that household income was positively associated with adult BMI tertiles. It was found that low-income families spent less than the estimated cost required for a basic diet. Furthermore, in opposition to the popular belief with respect to eating behaviours in low-income groups, these families only spent 5% of the total food expenditure on ready-to-eat food [Smith et al., 2013, Smith et al., 2010]. This might suggest that adults from low-income families still experienced food insecurity which might have protected them from obesity. However, I did not find a statistically significant effect of household income on overweight/obesity in children. The Working for Family program, which pays extra money to almost all families with children who earned less than \$57,000 per year, may explain the non-significant finding in children. This program ensured that low-income families with children would have enough money to afford food and their basic daily needs to sustain life. Further research on food securities among low-income families is warranted to shed more light on this issue, especially among those who do not receive any government benefit.

5.4 Obesity Trend

According to the OECD report in 2014, New Zealand is ranked third for its obesity prevalence; behind The United States and Mexico [OECD, 2017]. However, the figures reported in the OECD report did not account for the

difference in ethnic group composition and it would have not provided a correct picture of the obesity epidemic. For example, 61% of the US population is White, 12% is Black, and 18% is Hispanic in 2016 [Kaiser Family Foundation, 2016]; and their obesity prevalence was 47.8% in Non-Hispanic Black and 42.5% in Hispanic, while it was only 32.6% among Non-Hispanic White [Ogden et al., 2013]. The figure presented in the OECD report for the US must have been influenced by the high rate of obesity among Non-Hispanic Black and Hispanic group. Similarly, in New Zealand, 74% of the population is European, 15% is Māori, and 7% is Pacific [Statistics New Zealand, 2017] in 2013; and this study found that 55% Māori and 74% Pacific adults were obese in 2012/13.

By contrast, the United Kingdom, which had a lower national obesity prevalence according to OECD report, consists of 86% White and 2.5% Indian [Office for National Statistics, 2011]. This would translate to a relatively lower national obesity prevalence compared with countries with a high proportion of ethnic groups who have predispositions to accumulate more body weight. Using universal obesity cut-offs across different ethnicities will result in lower obesity prevalence among Asian and White European people while it will be higher for Black African [El-Sayed et al., 2011], Non-Hispanic Black, Hispanic [Ogden et al., 2013], Māori, and Pacific people. Comparing the obesity prevalence across countries while disregarding each country's unique ethnic composition will not give an accurate picture of the problem. It is better to report obesity prevalence on each ethnic group when trying to make a comparison of the obesity epidemic across different countries.

5.5 Ethnic Differences and The Universal BMI Cut-offs

The effect of ethnicities on BMI tertiles or categories persisted even after controlling for other demographic variables. In this study, Asians (Indian and/or Chinese) were less likely to be in the higher BMI tertiles compared with Europeans. This finding agrees with other studies [Stommel and Schoenborn, 2010, El-Sayed et al., 2011, Barba et al., 2004], and the trend is also observed in other Asian ethnic groups such as Indians, Chinese, Malaysians [Deurenberg-Yap et al., 2002, Deurenberg-Yap et al., 2000], Indonesians, Japanese [Barba et al., 2004], and Polynesians [Deurenberg, 2001]. Despite the tendency of having a lower average BMI, Asians have a higher risk of developing type II diabetes and cardiovascular disease compared with Europeans at the same BMI level. On average, South Asians are 3.4 times more likely to develop diabetes compared with Europeans after adjusting for age, sex, baseline BMI and other sociodemographic factors [Chiu et al., 2011]. This tells us that a higher BMI level do not always mean a higher risk of developing diabetes when looking at a population with diverse ethnic composition.

Compared with Caucasians at the same BMI level; Indians, Chinese, and Malays have a higher body fat percentage [Deurenberg-Yap et al., 2002, Deurenberg-Yap et al., 2000]. Despite that, Indians were still more likely to have insulin resistance compared with Caucasians even after controlling for diet, anthropometric measurement,

and body fat percentage [Abate and Chandalia, 2001]. Other ethnic groups such as American Indians, African American, Pacific Islanders [Dabelea et al., 2007], Chinese, Japanese, Malaysians also had a higher incidence of insulin resistance and type II diabetes compared with Europeans [Gao et al., 2012]. It may be possible that Asians may have a genetic predisposition to develop insulin resistance [Abate and Chandalia, 2001]. This association was further supported by a cohort study in Canada. It found that some ethnic groups develop type II diabetes at a much younger age than those of European descents after controlling for demographic variables and BMI. On average, diabetes presented itself nine years earlier in South Asian descent, three years earlier in Chinese descent, and one year earlier among Black descent compared with European descents [Chiu et al., 2011]. These suggest that BMI is a poor predictor of diabetes risk across diverse ethnic groups and there may be a genetic influence at play.

Despite the fact that Pacific and Māori people had a lower body fat percentage compared with other ethnic groups at the same BMI level [Rush et al., 2009], diabetes prevalence among Pacific and Māori remained high relative to the rest of New Zealand population [Hawley and McGarvey, 2015]. This study also found that Pacific adults were eight times and Pacific children were almost four times more likely to be in the higher BMI groups compared with Europeans. The same association was also found in Māori, where Māori adults and children were three and two times more likely to be in the higher BMI groups compared with Europeans. Furthermore, by using the universal BMI cut-offs, more than 70% of Pacific adults and more than half of Māori adults would be identified as obese. If overweight status was included, more than 90% and 80% of Pacific and Māori adults would be identified as either overweight or obese. This further questions the usefulness of the "average" BMI category and more research is needed in order to determine whether existing overweight/obesity cut-offs are accurate in predicting diabetes risk in these groups.

5.6 Diet and Physical Activity

5.6.1 Fruit and Vegetable Guidelines

I found no evidence that adherence to fruit and vegetable guidelines influence the BMI outcome after adjusting for deprivation level. A systematic review have found that a reduction in overweight/obesity likelihood is only achieved when the increase in fruit and vegetable consumption is accompanied by a reduction in the total caloric intake [Ledoux et al., 2011]. Another longitudinal study in children also found that fruit and vegetable consumption was not associated with less energy-dense food consumption (e.g. high-energy drinks, sweets, and snacks); which meant the children would have the same, if not more, total caloric intake than those who consumed less fruit and vegetable [Bayer et al., 2014]. The result from this study may had happened because people who did consume more fruit and vegetables had the same total caloric intake as those who did not.

5.6.2 Fast Food and Soft Drink Consumption in Children

I found that the effect of soft drink consumption on childhood overweight/obesity remained significant even after adjusting for parent's educational qualification, age, ethnicity, and deprivation level. The association between soft drink or SSB consumption on weight gain is consistent across studies [Vartanian et al., 2007, De Ruyter et al., 2012] even after adjusting for various diets and lifestyle factors [Malik et al., 2013a]. Interestingly, the association was attenuated when total energy intake was taken into account. Even so, soft drink consumption is consistently associated with increased total energy intake. This suggests that people who consumed SSB might not have been able to compensate for the extra energy [Vartanian et al., 2007]. Following administration of 500ml of either water or SSB, people would still consume the same amount of energy from food. This would result in a higher total energy intake for those who drank SSB compared with those who only drank water [Maersk et al., 2012]. Even though the effect of SSB consumption is attenuated by the total caloric intake, it is clear that those who drank SSB did not reduce their total caloric intake, this consequently put them in a higher risk of being obese.

On the other hand, I found that the frequency of fast food consumption was not associated with overweight/obesity status in children. This result agrees with a study in China regarding child obesity and fast food consumption. Even though there was a significant increase in fast food consumption among adolescents from 2004 to 2009; it was not accompanied by any change in BMI z-score after adjusting for age, ethnicity, household income, geographical region, and physical activity level [Xue et al., 2016]. Overall, the associations between fast food consumption and obesity were not found in children, but it is consistently significant in adults. A Possible explanation for this is that children require more energy while growing up, therefore extra caloric intake from fast food is not readily converted into body fat [Rosenheck, 2008]. Despite that, because fast food consumption was associated with soft drink consumption, it is possible that children who consumed fast food would increase their soft drink intake and in turn lead to a higher likelihood of being obese. The non-significant finding from this study might be a consequence of multi-collinearity between soft drink and fast food consumption.

5.6.3 Physical Activity

This study found that sedentary lifestyle (spending less than 30 minutes of exercise in the past week) did not have a significant influence on the BMI tertiles after adjusting for difficulty climbing several flights of stairs, which is a proxy for physical capability. It is possible that people who were sedentary had medical conditions that might lower their capability to exercise and make them more prone to developing obesity. For example, obese people have a higher risk of developing knee problems [Banjare and Bhalerao, 2016], which in turn lead to an increase in sedentary activity [Joffe, 2016] and make it difficult for them to reverse their obesity. On the other hand, people who were physically active (spent at least 30 minutes of exercise on five or more days in the past week) had a

lower likelihood of being in the higher BMI tertiles. Nevertheless, because reverse causation cannot be ruled out in this type of study and the evidence have shown that physical activity intervention was not effective in maintaining long-term weight loss [Kovacs et al., 2014, Shaw et al., 2006], it is unlikely that being physically active would reduce the risk of obesity. Instead, this finding may indicate that people with less functional limitations, which had normal weight, were more likely to be physically active.

5.7 Smoking Status

This study found that current smokers had a lower likelihood of being in the higher BMI tertiles compared with the non-smokers. In contrast, ex-smokers had a higher risk of being in the higher BMI tertiles. These findings are consistent with other studies where people tend to gain more weight after smoking cessation [MacKay et al., 2013, Rom et al., 2015, Siahpush et al., 2014]. The reduction in body weight among smokers is expected because nicotine activates the sympathetic nervous system and increases energy expenditure [Rigotti and Clair, 2013]. The weight gain after smoking cessation is observed as early as one year after quitting, but only in older adults. However, the weight gain was not observed in people aged 16–24 years who quit smoking [MacKay et al., 2013]. Although the weight gained is mostly attributable to an increase in body fat; a study showed that ex-smokers also had higher muscle mass, muscle strength, and bone density compared with current smokers [Rom et al., 2015]. Additionally, overweight or obese ex-smokers have a lower overall mortality risk compared with current smokers with normal weight [Siahpush et al., 2014]. Therefore, there is no benefit in preventing people who want to quit smoking in order to stay lean because continuing smoking even with normal weight would put people in a higher risk of developing adverse health events. It is noteworthy that an intervention focusing on one aspect of health (i.e. body weight) may not be appropriate for some people and careful considerations are needed when engaging overweight/obese individuals with unique problems.

5.8 Healthy Immigrant

This study found that adults who had always stayed in New Zealand and those who had migrated to New Zealand for more than 10–11 years ago were more likely to be in the higher BMI tertiles compared with those who migrated to New Zealand less than 10–11 years ago. The protective effect of being a migrant remained significant regardless of their ethnicity, age, education level, and sex. This finding is similar to studies from other countries such as the US, Canada, Australia, and the United Kingdom [Vang et al., 2015, Kennedy et al., 2015, Goel et al., 2004]. Better BMI outcome among the immigrants might be explained by the immigration selection favoring healthy individuals [Kennedy et al., 2015]. However, it is very likely that the protective effect will diminish over time, and the health risk of immigrants will match the risk of the host country after 10–15 years [Vang et al., 2015, Ledoux

et al., 2011, Goel et al., 2004].

Interestingly, another study did not find the healthy immigrant effect among Pacific people in New Zealand. It was found that not only Pacific people had received preference over other ethnic groups, but they also came from a more health disadvantaged communities [Hajat et al., 2010]. This study did not conduct a sub-group analysis, thus could not draw a specific conclusion regarding healthy immigrant effect among Pacific people. In light of that, interpreting the healthy immigrant effect should be done carefully as it might not apply to some groups of people due to other circumstances that were not captured in this study.

5.9 Urbanisation

Urbanisation has been thought to cause the obesity epidemic through various mechanisms [Kearney, 2010, Malik et al., 2013b, Sturm and An, 2014, Durand et al., 2011], but the evidence is conflicting. A study in India found that rural people who migrated to urban areas had a higher BMI and diabetes prevalence compared with rural dwellers. Those who migrated to urban areas had an increased fat intake and reduced physical activity, which may explain higher levels of obesity and diabetes among this group [Ebrahim et al., 2010]. However, a study in China found that people who lived in rural areas were more likely to be overweight/obese compared with those in urban areas. It appears that China has managed to improve food security even in rural settings, which lead to higher caloric consumption. Moreover, due to agricultural mechanisation, a lot more Chinese people in rural areas did not exercise as often as people in metropolitan areas [Tian et al., 2014]. This study found that there was no BMI difference between urban and rural dwellers. This finding suggests that the urban/rural indicator may not be a reliable predictor of overweight/obesity status. There may be more important demographic information such as deprivation level and ethnicity, which have a stronger influence on BMI.

5.10 Study Strengths and Limitations

To my knowledge, this is the first study in New Zealand that uses panels of cross-sectional surveys of individual data to assess the association between deprivation level and BMI. The analysis was done using the NZHS data, which contains information on various demographic variables and health-related behaviours. In that regard, the association between deprivation level and BMI outcome in this study is reflective of the true effect estimate as it is adjusted for possible confounding factors such as sex, age, household income, smoking status, physical activity level, and diets. This study also used a complex survey design, which took into account the sampling design used in the NZHSs. This ensured the result of this analysis is generalisable to the New Zealand population.

Despite the use of multiple cross-sectional surveys over time, findings in this study is not free from the possibilities of reverse causation. This study also did not have access to individual identifiers, and thus are unable to

exclude the possibility that the same person could be measured more than once throughout the survey periods. Moreover, the information on deprivation quintile in NZHS 2011/12 and 2012/13 came from the census data in 2006; this study assumed that the area level deprivation had not changed significantly as to have an impact on the BMI outcome at those survey periods. The NZHS also did not have information on migrations within New Zealand, it is possible that healthier people moved into communities with a lower deprivation level which might have confounded the findings. This study also did not conduct subgroup analyses across ethnic groups. Considering that more than 80% of Pacific and Māori belong to the higher BMI categories, it is possible that the relationship between area level deprivation and the BMI outcomes might change for a certain group.

6 Conclusions

6.1 Implications of Socio-Economic Status and BMI outcomes

Higher levels of area deprivation put people at a higher risk of developing obesity. This finding persists even after adjusting for other demographic factors and health-related behaviours. Lack of socio-economic and health care resources in this area makes it harder for people to prevent and/or reverse obesity. Imposing behavioural changes on these people should be done with consideration, as they might not have the resources (i.e. time, money, facilities, etc.) to modify their habits. In order to reduce the disparity in the obesity prevalence across different area deprivation levels, the government should devise a population-level intervention that improves the socio-economic status of the area and also create an environment that is conducive to maintaining a healthy lifestyle. This study also found that, unlike other high-income countries, high-income families in New Zealand had a higher risk of developing obesity. This may indicate that low-income families still experienced food security issues in their everyday lives.

6.2 Implications of The Universal BMI Cut-Offs

The intention of categorising people into different BMI groups is to identify the health risks associated with a certain BMI group and to devise appropriate interventions that may mitigate the associated health risks. However, as I have discussed, universal BMI cut-offs are not accurate in assessing the obesity-associated health risks in a population with diverse ethnic groups. Different ethnic groups have different average body weight, and their risk of developing diabetes is different at the same BMI value. We found that the obesity prevalence in Asians was just under 15% and despite having a much lower average BMI, their risk of developing diabetes is higher compared with the other ethnic groups at the same BMI level. Using BMI alone as an indicator of population health would mean that Asians will receive little attention. On the other hand, a higher average BMI among Pacific and Māori would mean that they will receive rigorous interventions which lead to another problem. If Pacific and Māori have a higher average BMI, at what BMI point can the intervention be said to have achieved its goal? Further imposing more strenuous diet and/or exercise interventions would only create frustration without achieving the actual goals (i.e. reducing the risk of obesity-associated comorbidities). Forcing the use of the "average" BMI category as a blanket public health goal may harm certain ethnic groups.

6.3 Implications of Diet and Physical Activity Interventions

This study found that lower soft drink consumption is associated with lower likelihood to be in the higher BMI categories. However, no association is found for fast food consumption as well as adherence to fruit and vegetable guidelines. This, however, does not mean consuming more fast food will not have any impact on BMI. Due to correlation between fast food and soft drink consumption, it is likely that people who consume more fast food will increase their soft drink intake as well; this in turn will increase the risk of developing obesity. Replacing fast food and soft drink with more fruit and vegetables, despite having no effect on BMI, would fulfill the daily nutritional and micro-nutrients requirement. It is important to note that improving population dietary habits with more fruit and vegetables will benefit health regardless of weight change, and no weight change should not be used as an argument to deter from eating healthier foods.

The result from this study also demonstrates that there was no difference in obesity risk between people who had a sedentary and non-sedentary lifestyle. The benefit of physical activity on reducing the likelihood of being in the higher BMI groups was only realised when people spent more than 30 minutes of moderate physical activity or brisk walking on at least five days in a week. This criteria appears to be unrealistic because only a small fraction of the population can achieve this. Moreover, it is possible that those who did not engage in physical activity might have had other medical problems that limit their functional capacity. Standard physical activity interventions may not be suitable for this group of people, and imposing a desired BMI on them is highly inappropriate. As have been shown from multiple studies, physical activity interventions did not manage to maintain long-term weight loss thus the drive to increase physical activity level should not be based on weight reduction. Increasing physical activity whenever and wherever appropriate will benefit other aspects of health (e.g. improved blood pressure, lower lipid profiles and fasting serum glucose) regardless of weight change.

6.4 Recommendations

People living in more deprived areas are faced with socio-economic barriers and structural challenges in achieving a healthier lifestyle. Therefore, devising public health interventions and/or health policies to reduce obesity should not only empower them with knowledge about nutrition and exercise but also improve infrastructure support (e.g., more accessible and healthier food options, better access to health care, etc.).

Furthermore, reliance on BMI or other body weight measurements as the only population or individual health goal should stop because different ethnic groups have a different risk of developing adverse health conditions at the same BMI level. Instead, dietary habits and physical activity levels before and after an intervention can be used as an indicator of successful progress towards a healthier life regardless of weight change.

Longitudinal research assessing the impact of area level deprivation on obesity-associated comorbidities in New Zealand is needed. This is important because obesity on its own is not the main problem, whereas the diseases associated with obesity are the conditions that impact on quality of life and increase health care costs. Also, more research is needed to assess the risk of developing obesity-associated chronic conditions among Pacific and Māori people in New Zealand. This will also help determine whether specific overweight/obesity cut-offs or a better indicator of population health other than weight measurements are needed for these groups.

Bibliography

- [Abate and Chandalia, 2001] Abate, N. and Chandalia, M. (2001). Ethnicity and type 2 diabetes: Focus on Asian Indians. *Journal of Diabetes and its Complications*, 15(6):320–327.
- [Abdullah et al., 2011] Abdullah, A., Stoelwinder, J., Shortreed, S., Wolfe, R., Stevenson, C., Walls, H., de Courten, M., and Peeters, A. (2011). The duration of obesity and the risk of type 2 diabetes. *Public Health Nutrition*, 14(1):119–126.
- [Agresti, 2013] Agresti, A. (2013). *Categorical data analysis*, volume 792. Wiley, Hoboken, NJ, 3rd edition.
- [Alexandratos and Bruinsma, 2012] Alexandratos, N. and Bruinsma, J. (2012). World agriculture towards 2030/2050: The 2012 revision. Report, ESA Working paper Rome, FAO.
- [Angkurawaranon et al., 2014] Angkurawaranon, C., Jiraporncharoen, W., Chenthanakij, B., Doyle, P., and Nitsch, D. (2014). Urban environments and obesity in Southeast Asia: a systematic review, meta-analysis and meta-regression. *PLoS ONE*, 9(11):1–19.
- [Antonini et al., 2007] Antonini, L., Grosso, A., Francalanci, C., and Mattei, R. (2007). Cardiovascular pathophysiology of obesity. *High Blood Pressure & Cardiovascular Prevention*, 14(4):207–212.
- [Antonopoulos et al., 2016] Antonopoulos, A. S., Oikonomou, E. K., Antoniadou, C., and Tousoulis, D. (2016). From the BMI paradox to the obesity paradox: the obesity-mortality association in coronary heart disease: obesity paradox in CHD. *Obesity Reviews*, 17(10):989–1000.
- [Atkinson et al., 2014] Atkinson, J., Salmond, C., and Crampton, P. (2014). *NZDep2013 Index of Deprivation*. University of Otago, Dunedin.
- [Banjare and Bhalerao, 2016] Banjare, J. B. and Bhalerao, S. (2016). Obesity associated noncommunicable disease burden. *International Journal of Health & Allied Sciences*, 5(2):81–87.
- [Barba et al., 2004] Barba, C., Cavalli-Sforza, T., Cutter, J., Darnton-Hill, I., Deurenberg, P., Deurenberg-Yap, M., Gill, T., James, P., Ko, G., Nishida, C., and Consultation, W. H. O. E. (2004). Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet*, 363(9403):157–163.
- [Barnes et al., 2007] Barnes, L. A., Opitz, J. M., and Gilbert-Barnes, E. (2007). Obesity: genetic, molecular, and environmental aspects. *American Journal of Medical Genetics Part A*, 143A(24):3016–3034.

- [Bayer et al., 2014] Bayer, O., Nehring, I., Bolte, G., von Kries, R., and Cohort, G. (2014). Fruit and vegetable consumption and BMI change in primary school-age children: A cohort study. *European Journal of Clinical Nutrition*, 68(2):265–270.
- [Bogardus, 2009] Bogardus, C. (2009). Missing heritability and GWAS utility. *Obesity*, 17(2):209–210.
- [Buchmueller and Johar, 2015] Buchmueller, T. C. and Johar, M. (2015). Obesity and health expenditures: Evidence from Australia. *Economics & Human Biology*, 17:42–58.
- [Buchwald et al., 2004] Buchwald, H., Avidor, Y., Braunwald, E., Jensen, M. D., Pories, W., Fahrbach, K., and Schoelles, K. (2004). Bariatric surgery: A systematic review and meta-analysis. *JAMA*, 292(14):1724–1737.
- [Cali and Caprio, 2008] Cali, A. M. G. and Caprio, S. (2008). Obesity in children and adolescents. *The Journal of Clinical Endocrinology & Metabolism*, 93(11_supplement_1):s31–s36.
- [Centers for Disease Control and Prevention, 2016] Centers for Disease Control and Prevention (16 June 2016). Defining adult overweight and obesity. <http://www.cdc.gov/obesity/adult/defining.html>. Accessed: 22 April 2017.
- [Centers for Disease Control and Prevention, 2013] Centers for Disease Control and Prevention (May 2013). Use and interpretation of the WHO and CDC growth chart for children from birth to 20 years in the united states. <https://www.cdc.gov/nccdphp/dnpa/growthcharts/resources/growthchart.pdf>. Accessed: 22 April 2017.
- [Chaparro et al., 2014] Chaparro, M. P., Whaley, S. E., Crespi, C. M., Kroleit, M., Nobari, T. Z., Seto, E., and Wang, M. C. (2014). Influences of the neighbourhood food environment on adiposity of low-income preschool-aged children in los angeles county: a longitudinal study. *Journal of Epidemiology and Community Health*, 68(11):1027–1033.
- [Chiu et al., 2011] Chiu, M., Austin, P. C., Manuel, D. G., Shah, B. R., and Tu, J. V. (2011). Deriving ethnic-specific BMI cutoff points for assessing diabetes risk. *Diabetes Care*, 34(8):1741–1748.
- [Cohen et al., 2013] Cohen, A. K., Rai, M., Rehkopf, D. H., and Abrams, B. (2013). Educational attainment and obesity: A systematic review. *Obesity Reviews*, 14(12):989–1005.
- [Cohen, 1992] Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1):155–159.
- [Cole et al., 2000] Cole, T. J., Bellizzi, M. C., Flegal, K. M., and Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: International survey. *British Medical Journal*, 320(7244):1240–1243.

- [Cole and Lobstein, 2012] Cole, T. J. and Lobstein, T. (2012). Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatric Obesity*, 7(4):284–294.
- [coreysparks, 2015] coreysparks (2015). DEM 7283 — Example 4 — ordinal & multinomial logit models. https://rstudio-pubs-static.s3.amazonaws.com/58926_5f62fd1d242745c4a2a28b18376118df.html. Accessed: 10 April 2017.
- [Cotti and Tefft, 2013] Cotti, C. and Tefft, N. (2013). Fast food prices, obesity, and the minimum wage. *Economics and Human Biology*, 11(2):134–147.
- [Craig et al., 2016] Craig, L. C. A., McNeill, G., Masson, L. F., and Macdiarmid, J. I. (2016). Diet quality index in children in scotland: associations with age, sex, socio-economic deprivation and obesity. *Proceedings of the Nutrition Society*, 75(OCE3).
- [Crombie et al., 2008] Crombie, I. K., Kiezebrink, K., Irvine, L., Wrieden, W. L., Swanson, V., Power, K., and Slane, P. W. (2009;2008;). What maternal factors influence the diet of 2-year-old children living in deprived areas? a cross-sectional survey. *Public Health Nutrition*, 12(8):1254–1260.
- [Cummins and Macintyre, 2006] Cummins, S. and Macintyre, S. (2006). Food environments and obesity — neighbourhood or nation? *International Journal of Epidemiology*, 35(1):100–104.
- [Dabelea et al., 2007] Dabelea, D., Bell, R. A., D’Agostino Jr, R. B., Imperatore, G., Johansen, J. M., Linder, B., Liu, L. L., Loots, B., Marcovina, S., Mayer-Davis, E. J., et al. (2007). Incidence of diabetes in youth in the united states. *JAMA*, 297(24):2716–2724.
- [De Ruyter et al., 2012] De Ruyter, J. C., Olthof, M. R., Seidell, J. C., and Katan, M. B. (2012). A trial of sugar-free or sugar-sweetened beverages and body weight in children. *New England Journal of Medicine*, 367(15):1397–1406.
- [de Vos et al., 2016] de Vos, B. C., Runhaar, J., van Middelkoop, M., Krul, M., and Bierma-Zeinstra, S. M. (2016). Long-term effects of a randomized, controlled, tailor-made weight-loss intervention in primary care on the health and lifestyle of overweight and obese women. *Am J Clin Nutr*, 104(1):33–40.
- [Deurenberg, 2001] Deurenberg, P. (2001). Universal cut-off BMI points for obesity are not appropriate. *British Journal of Nutrition*, 85(2):135–136.
- [Deurenberg-Yap et al., 2002] Deurenberg-Yap, M., Chew, S., and Deurenberg, P. (2002). Elevated body fat percentage and cardiovascular risks at low body mass index levels among singaporean chinese, malays and indians. *Obesity Reviews*, 3(3):209–215.

- [Deurenberg-Yap et al., 2000] Deurenberg-Yap, M., Schmidt, G., Staveren, van, W., and Deurenberg, P. (2000). The paradox of low body mass index and high body fat percentage among Chinese, Malays and Indians in singapore. *International Journal of Obesity*, 24(8):1011–1017.
- [Didham and Callister, 2012] Didham, R. and Callister, P. (2012). The effect of ethnic prioritisation on ethnic health analysis: a research note. *The New Zealand medical journal*, 125(1359):58–66.
- [Dijkstra et al., 2015] Dijkstra, S. C., Neter, J. E., van Stralen, M. M., Knol, D. L., Brouwer, I. A., Huisman, M., and Visser, M. (2015). The role of perceived barriers in explaining socio-economic status differences in adherence to the fruit, vegetable and fish guidelines in older adults: a mediation study. *Public Health Nutrition*, 18(5):797–808.
- [Dinsa et al., 2012] Dinsa, G. D., Goryakin, Y., Fumagalli, E., and Suhrcke, M. (2012). Obesity and socioeconomic status in developing countries: a systematic review. *Obesity Reviews*, 13(11):1067–1079.
- [Dugas et al., 2011] Dugas, L. R., Harders, R., Merrill, S., Ebersole, K., Shoham, D. A., Rush, E. C., Assah, F. K., Forrester, T., Durazo-Arvizu, R. A., and Luke, A. (2011). Energy expenditure in adults living in developing compared with industrialized countries: a meta-analysis of doubly labeled water studies. *The American journal of clinical nutrition*, 93(2):427–441.
- [Durand et al., 2011] Durand, C. P., Andalib, M., Dunton, G. F., Wolch, J., and Pentz, M. A. (2011). A systematic review of built environment factors related to physical activity and obesity risk: implications for smart growth urban planning. *Obesity Reviews*, 12(501):e173–e182.
- [Eagle et al., 2012] Eagle, T. F., Sheetz, A., Gurm, R., Woodward, A. C., Kline-Rogers, E., Leibowitz, R., Durussel-Weston, J., Palma-Davis, L., Aaronson, S., Fitzgerald, C. M., Mitchell, L. R., Rogers, B., Bruenger, P., Skala, K. A., Goldberg, C., Jackson, E. A., Erickson, S. R., and Eagle, K. A. (2012). Understanding childhood obesity in america: Linkages between household income, community resources, and children’s behaviors. *American Heart Journal*, 163(5):836–843.
- [Ebrahim et al., 2010] Ebrahim, S., Kinra, S., Bowen, L., Andersen, E., Ben-Shlomo, Y., Lyngdoh, T., Ramakrishnan, L., Ahuja, R. C., Joshi, P., Das, S. M., Mohan, M., Smith, G. D., Prabhakaran, D., Srinath Reddy, K., Grp, I. M. S., group, I. M. S., and for the Indian Migration Study group (2010). The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. *PLoS Medicine*, 7(4):e1000268.
- [El-Sayed et al., 2011] El-Sayed, A. M., Scarborough, P., and Galea, S. (2011). Ethnic inequalities in obesity among children and adults in the UK: a systematic review of the literature. *Obesity Reviews*, 12(501):e516–e534.

- [Evans et al., 2000] Evans, J. M. M., Newton, R. W., Ruta, D. A., MacDonald, T. M., and Morris, A. D. (2000). Socio-economic status, obesity and prevalence of type 1 and type 2 diabetes mellitus. *Diabetic Medicine*, 17(6):478–480.
- [Exeter et al., 2015] Exeter, D., Sabel, C., Hanham, G., Lee, A., and Wells, S. (2015). Movers and stayers: the geography of residential mobility and CVD hospitalisations in Auckland, New Zealand. *Social Science & Medicine*, 133:331–339.
- [Flegal et al., 2013] Flegal, K. M., Kit, B. K., Orpana, H., and Graubard, B. I. (2013). Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA*, 309(1):71–82.
- [Fruhstorfer et al., 2016] Fruhstorfer, B. H., Mousoulis, C., Uthman, O. A., and Robertson, W. (2016). Socio-economic status and overweight or obesity among school-age children in sub-saharan africa — a systematic review: obesity among children in sub-Saharan Africa. *Clinical Obesity*, 6(1):19–32.
- [Gao et al., 2012] Gao, H., Salim, A., Lee, J., Tai, E. S., and Van Dam, R. M. (2012). Can body fat distribution, adiponectin levels and inflammation explain differences in insulin resistance between ethnic Chinese, Malays and Asian Indians. *International Journal of Obesity*, 36(8):1086–1093.
- [Giskes et al., 2011] Giskes, K., van Lenthe, F., Avendano-Pabon, M., and Brug, J. (2011). A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obesity reviews*, 12(501):e95–e106.
- [Goel et al., 2004] Goel, M. S., McCarthy, E. P., Phillips, R. S., and Wee, C. C. (2004). Obesity among US immigrant subgroups by duration of residence. *Jama*, 292(23):2860–2867.
- [Guh et al., 2009] Guh, D. P., Zhang, W., Bansback, N., Amarsi, Z., Birmingham, C. L., and Anis, A. H. (2009). The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health*, 9(1):88–88.
- [Hajat et al., 2010] Hajat, A., Blakely, T., Dayal, S., and Jatrana, S. (2010). Do New Zealand’s immigrants have a mortality advantage? evidence from the New Zealand Census-Mortality Study. *Ethnicity & Health*, 15(5):531–547.
- [Hawley and McGarvey, 2015] Hawley, N. L. and McGarvey, S. T. (2015). Obesity and diabetes in Pacific Islanders: the current burden and the need for urgent action. *Current Diabetes Reports*, 15(5):1–10.
- [Herrera and Lindgren, 2010] Herrera, B. M. and Lindgren, C. M. (2010). The genetics of obesity. *Current Diabetes Reports*, 10(6):498–505.

- [Hodgkin et al., 2010] Hodgkin, E., Hamlin, M. J., Ross, J. J., and Peters, F. (2010). Obesity, energy intake and physical activity in rural and urban New Zealand children. *Rural and remote health*, 10(2):1336.
- [Hu et al., 2014] Hu, Y., Bhupathiraju, S. N., Koning, L., and Hu, F. B. (2014). Duration of obesity and overweight and risk of type 2 diabetes among US women. *Obesity*, 22(10):2267–2273.
- [Huxley et al., 2010] Huxley, R., Mendis, S., Zheleznyakov, E., Reddy, S., and Chan, J. (2010). Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk: a review of the literature. *European Journal of Clinical Nutrition*, 64(1):16–22.
- [Jensen et al., 2013] Jensen, M. D., Ryan, D. H., Apovian, C. M., Ard, J. D., Comuzzie, A. G., Donato, K. A., Hu, F. B., Hubbard, V. S., Jakicic, J. M., Kushner, R. F., et al. (2013). 2013 aha/acc/tos guideline for the management of overweight and obesity in adults. *Circulation*, pages 01–cir.
- [Joffe, 2016] Joffe, A. (2016). Adolescent obesity and functional limitations in young adulthood. *NEJM Journal Watch: Pediatrics & Adolescent Medicine*.
- [Johnson and Wardle, 2011] Johnson, F. and Wardle, J. (2011). Socio-economic status and obesity: epidemiology and explanations. *Psychologist*, 24(12):893–893.
- [Johnson and Johnson, 2015] Johnson, J. A. and Johnson, A. M. (2015). Urban-rural differences in childhood and adolescent obesity in the United States: A systematic review and meta-analysis. *Childhood Obesity*, 11(3):233–241.
- [Jull et al., 2011] Jull, A., Lawes, C. M. M., Eyles, H., Maddison, R., Gorton, D., Arcus, K., Chee, N., Taylor, B., and Mann, J. (2011). Clinical guidelines for weight management in new zealand adults, children and young people. *Journal of Primary Health Care*, 3(1):66–71.
- [Juonala et al., 2011] Juonala, M., Magnussen, C. G., Berenson, G. S., Venn, A., Burns, T. L., Sabin, M. A., Srinivasan, S. R., Daniels, S. R., Davis, P. H., Chen, W., Sun, C., Cheung, M., Viikari, J. S. A., Dwyer, T., and Raitakari, O. T. (2011). Childhood adiposity, adult adiposity, and cardiovascular risk factors. *New England Journal of Medicine*, 365(20):1876–1885.
- [Kaiser Family Foundation, 2016] Kaiser Family Foundation (2016). Population distribution by race/ethnicity. <http://kff.org/other/state-indicator/distribution-by-raceethnicity/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>. Accessed: 5 June 2017.
- [Kearney, 2010] Kearney, J. (2010). Food consumption trends and drivers. *Philosophical Transactions: biological Sciences*, 365(1554):2793–2807.

- [Kelly and Swinburn, 2015] Kelly, S. and Swinburn, B. (2015). Childhood obesity in New Zealand. *The New Zealand medical journal*, 128(1417):6–7.
- [Kennedy et al., 2015] Kennedy, S., Kidd, M. P., McDonald, J. T., and Biddle, N. (2015). The healthy immigrant effect: patterns and evidence from four countries. *Journal of International Migration and Integration*, 16(2):317–332.
- [Kovacs et al., 2014] Kovacs, E., Siani, A., Konstabel, K., Hadjigeorgiou, C., de Bourdeaudhuij, I., Eiben, G., Lissner, L., Gwozdz, W., Reisch, L., Pala, V., Moreno, L. A., Pigeot, I., Pohlmann, H., Ahrens, W., and Molnar, D. (2014). Adherence to the obesity-related lifestyle intervention targets in the IDEFICS study. *International Journal of Obesity*, 38(S2):S144–S151.
- [Lai et al., 2014] Lai, C., Sun, D., Cen, R., Wang, J., Li, S., Fernandez-Alonso, C., Chen, W., Srinivasan, S. R., and Berenson, G. S. (2014). Impact of long-term burden of excessive adiposity and elevated blood pressure from childhood on adulthood left ventricular remodeling patterns: the Bogalusa Heart Study. *Journal of the American College of Cardiology*, 64(15):1580–1587.
- [Le et al., 2016] Le, H., Engler-Stringer, R., and Muhajarine, N. (2016). Walkable home neighbourhood food environment and children’s overweight and obesity: proximity, density or price? *Canadian Journal of Public Health*, 107(1):ES42–ES47.
- [Ledoux et al., 2011] Ledoux, T. A., Hingle, M. D., and Baranowski, T. (2011). Relationship of fruit and vegetable intake with adiposity: a systematic review. *Obesity Reviews*, 12(501):e143–e150.
- [Li et al., 2008] Li, G., Li, H., Li, H., Zhang, P., Zhang, J., Zhang, B., Wang, J., Gregg, E. W., Yang, W., Gong, Q., Jiang, Y., An, Y., Shuai, Y., Thompson, T. J., Gerzoff, R. B., Roglic, G., Hu, Y., and Bennett, P. H. (2008). The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *The Lancet*, 371(9626):1783–1789.
- [Li et al., 2014] Li, X., Memarian, E., Sundquist, J., Zöller, B., and Sundquist, K. (2014). Neighbourhood deprivation, individual-level familial and socio-demographic factors and diagnosed childhood obesity: a nationwide multilevel study from Sweden. *Obesity facts*, 7(4):253–263.
- [Luke and Cooper, 2013] Luke, A. and Cooper, R. S. (2013). Physical activity does not influence obesity risk: time to clarify the public health message. *International Journal of Epidemiology*, 42(6):1831–1836.
- [Luke et al., 2011] Luke, A., Dugas, L. R., Durazo-Arvizu, R. A., Cao, G., and Cooper, R. S. (2011). Assessing physical activity and its relationship to cardiovascular risk factors: Nhanes 2003-2006. *BMC Public Health*, 11(1):387–387.

- [Lumley, 2004] Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9(1):1–19. R package version 2.2.
- [Lumley, 2016] Lumley, T. (2016). Survey: analysis of complex survey samples. R package version 3.31-5.
- [MacKay et al., 2013] MacKay, D. F., Gray, L., and Pell, J. P. (2013). Impact of smoking and smoking cessation on overweight and obesity: Scotland-wide, cross-sectional study on 40,036 participants. *BMC Public Health*, 13(1):348–348.
- [Maersk et al., 2012] Maersk, M., Belza, A., Holst, J. J., Fenger-Grøn, M., Pedersen, S. B., Astrup, A., and Richelsen, B. (2012). Satiety scores and satiety hormone response after sucrose-sweetened soft drink compared with isocaloric semi-skimmed milk and with non-caloric soft drink: a controlled trial. *European Journal of Clinical Nutrition*, 66(4):523–529.
- [Maier et al., 2013] Maier, W., Holle, R., Hunger, M., Peters, A., Meisinger, C., Greiser, K., Kluttig, A., Völzke, H., Schipf, S., Moebus, S., et al. (2013). The impact of regional deprivation and individual socio-economic status on the prevalence of type 2 diabetes in germany. a pooled analysis of five population-based studies. *Diabetic Medicine*, 30(3).
- [Malik et al., 2013a] Malik, V. S., Pan, A., Willett, W. C., and Hu, F. B. (2013a). Sugar-sweetened beverages and weight gain in children and adults: A systematic review and meta-analysis. *American Journal of Clinical Nutrition*, 98(4):1084–1102.
- [Malik et al., 2010] Malik, V. S., Popkin, B. M., Bray, G. A., Després, J.-P., Willett, W. C., and Hu, F. B. (2010). Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes. *Diabetes Care*, 33(11):2477–2483.
- [Malik et al., 2013b] Malik, V. S., Willett, W. C., and Hu, F. B. (2013b). Global obesity: trends, risk factors and policy implications. *Nature Reviews Endocrinology*, 9(1):13–27.
- [Matheson et al., 2008] Matheson, F. I., Moineddin, R., and Glazier, R. H. (2008). The weight of place: A multilevel analysis of gender, neighborhood material deprivation, and body mass index among canadian adults. *Social Science & Medicine*, 66(3):675–690.
- [Mattes and Foster, 2014] Mattes, R. and Foster, G. D. (2014). Food environment and obesity. *Obesity*, 22(12):2459–2461.
- [McCormack and Shiell, 2011] McCormack, G. R. and Shiell, A. (2011). In search of causality: A systematic review of the relationship between the built environment and physical activity among adults. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1):125–125.

- [Mei et al., 2012] Mei, H., Chen, W., Jiang, F., He, J., Srinivasan, S., Smith, E. N., Schork, N., Murray, S., and Berenson, G. S. (2012). Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the course of childhood and adulthood. *PLoS ONE*, 7(2):e31470.
- [Min et al., 2013] Min, J., Chiu, D. T., and Wang, Y. (2013). Variation in the heritability of body mass index based on diverse twin studies: a systematic review. *Obesity Reviews*, 14(11):871–882.
- [Ministry of Health, 2009] Ministry of Health (2009). *Report on New Zealand cost-of-illness studies on long-term conditions*. Ministry of Health, Wellington.
- [Ministry of Health, 2015a] Ministry of Health (2015a). *Annual update of key results 2014/15: New Zealand health survey*. Ministry of Health, Wellington.
- [Ministry of Health, 2015b] Ministry of Health (2015b). *Methodology report for the 2014/15 New Zealand Health Survey*. Ministry of Health, Wellington.
- [Ministry of Health, 2016] Ministry of Health (2016). *Annual update of key results 2015/16: New Zealand health survey*. Ministry of Health, Wellington.
- [Morrissey et al., 2014] Morrissey, T. W., Jacknowitz, A., and Vinopal, K. (2014). Local food prices and their associations with children’s weight and food security. *Pediatrics*, 133(3):422–430.
- [Nau et al., 2015] Nau, C., Schwartz, B. S., Bandeen-Roche, K., Liu, A., Pollak, J., Hirsch, A., Bailey-Davis, L., and Glass, T. A. (2015). Community socioeconomic deprivation and obesity trajectories in children using electronic health records. *Obesity*, 23(1):207–212.
- [Neff et al., 2013] Neff, K. J., Olbers, T., and le Roux, C. W. (2013). Bariatric surgery: the challenges with candidate selection, individualizing treatment and clinical outcomes. *BMC Medicine*, 11(1):8.
- [Newton et al., 2017] Newton, S., Braithwaite, D., and Akinyemiju, T. (2017). Socio-economic status over the life course and obesity: systematic review and meta-analysis. *PLoS ONE*, 12(5).
- [Ni Mhurchu and Ogra, 2007] Ni Mhurchu, C. and Ogra, S. (2007). The price of healthy eating: cost and nutrient value of selected regular and healthier supermarket foods in New Zealand. *The New Zealand Medical Journal*, 120(1248):U2388.
- [Norredam et al., 2014] Norredam, M., Agyemang, C., Hansen, H., Oluf, K., Petersen, J. H., Byberg, S., Krasnik, A., and Kunst, A. E. (2014). Duration of residence and disease occurrence among refugees and family reunited immigrants: test of the “healthy migrant” effect hypothesis. *Tropical Medicine & International Health*, 19(8):958–967.

- [OECD, 2017] OECD (2017). Obesity update 2017. <http://www.oecd.org/els/health-systems/Obesity-Update-2017.pdf>. Accessed: 11 July 2017.
- [Office for National Statistics, 2011] Office for National Statistics (2011). Ethnicity and national identity in England and Wales: 2011. <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11>. Accessed: 22 June 2017.
- [Ogden et al., 2013] Ogden, C. L., Carroll, M. D., Kit, B. K., and Flegal, K. M. (2013). Prevalence of obesity among adults: United States, 2011-2012. *NCHS data brief*, (131):1.
- [O’Rahilly and Farooqi, 2008] O’Rahilly, S. and Farooqi, I. S. (2008). Human obesity: a heritable neurobehavioral disorder that is highly sensitive to environmental conditions. *Diabetes*, 57(11):2905–2910.
- [Padwal et al., 2003] Padwal, R. S., Rucker, D., Li, S. K., Curioni, C., and Lau, D. C. W. (2003). Long-term pharmacotherapy for obesity and overweight. *The Cochrane Library*.
- [Pedrosa et al., 2011] Pedrosa, C., Correia, F., Seabra, D., Oliveira, B. M., Simoes-Pereira, C., and Vaz-de-Almeida, M. D. (2011). Prevalence of overweight and obesity among 7-9-year-old children in Aveiro, Portugal: comparison between IOTF and CDC references. *Public Health Nutrition*, 14(01):14–19.
- [Pereira et al., 2005] Pereira, M. A., Kartashov, A. I., Ebbeling, C. B., Van Horn, L., Slattery, M. L., Jacobs, D. R., and Ludwig, D. S. (2005). Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. *The lancet*, 365(9453):36–42.
- [Powell et al., 2009] Powell, L. M., Zhao, Z., and Wang, Y. (2009). Food prices and fruit and vegetable consumption among young American adults. *Health and Place*, 15(4):1064–1070.
- [Rigotti and Clair, 2013] Rigotti, N. A. and Clair, C. (2013). Managing tobacco use: the neglected cardiovascular disease risk factor. *European Heart Journal*, 34(42):3259–3267.
- [Rom et al., 2015] Rom, O., Reznick, A. Z., Keidar, Z., Karkabi, K., and Aizenbud, D. (2015). Smoking cessation-related weight gain-beneficial effects on muscle mass, strength and bone health. *Addiction*, 110(2):326–335.
- [Romero-Corral et al., 2006] Romero-Corral, A., Montori, V. M., Somers, V. K., Korinek, J., Thomas, R. J., Allison, T. G., Mookadam, F., and Lopez-Jimenez, F. (2006). Association of bodyweight with total mortality and with cardiovascular events in coronary artery disease: a systematic review of cohort studies. *The Lancet*, 368(9536):666–678.

- [Rosenheck, 2008] Rosenheck, R. (2008). Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk. *Obesity Reviews*, 9(6):535–547.
- [Rush et al., 2009] Rush, E. C., Freitas, I., and Plank, L. D. (2009). Body size, body composition and fat distribution: comparative analysis of European, Maori, Pacific Island and Asian Indian adults. *British Journal of Nutrition*, 102(4):632–641.
- [Shaw et al., 2006] Shaw, K. A., Gennat, H. C., O'Rourke, P., and Mar, C. D. (2006). Exercise for overweight or obesity. *The Cochrane Library*.
- [Sheperd, 2005] Sheperd, T. M. (2005). Bariatric surgery in context. *Journal of Family Practice*, 54(2):3–9.
- [Siahpush et al., 2014] Siahpush, M., Singh, G., Tibbits, M., Pinard, C., Shaikh, R., and Yarocho, A. (2014). It is better to be a fat ex-smoker than a thin smoker: findings from the 1997-2004 national health interview survey-national death index linkage study. *Tobacco Control*, 23(5):395–402.
- [Singh et al., 2015] Singh, G. K., Siahpush, M., Azuine, R. E., and Williams, S. D. (2015). Increasing area deprivation and socioeconomic inequalities in heart disease, stroke, and cardiovascular disease mortality among working age populations, united states, 1969-2011. *International Journal of MCH and AIDS*, 3(2):119–133.
- [Sjöström et al., 2012] Sjöström, L., Peltonen, M., Jacobson, P., Sjöström, C. D., Karason, K., Wedel, H., Ahlin, S., Anveden, Å., Bengtsson, C., Bergmark, G., et al. (2012). Bariatric surgery and long-term cardiovascular events. *JAMA*, 307(1):56–65.
- [Smith et al., 2010] Smith, C., Parnell, W., and Brown, R. (2010). *Family food environment: barriers to acquiring affordable and nutritious food in New Zealand households*. Families Commission.
- [Smith et al., 2013] Smith, C., Parnell, W. R., Brown, R. C., and Gray, A. R. (2013). Balancing the diet and the budget: food purchasing practices of food-insecure families in New Zealand. *Nutrition & Dietetics*, 70(4):278–285.
- [Stafford et al., 2010] Stafford, M., Brunner, E. J., Head, J., and Ross, N. A. (2010). Deprivation and the development of obesity. *American Journal of Preventive Medicine*, 39(2):130–139.
- [Statistics New Zealand, 2017] Statistics New Zealand (2017). Ethnic group (total responses) by age group and sex, for the census usually resident population count, 2001, 2006, and 2013 censuses (RC, TA). <http://nzdotstat.stats.govt.nz/wbos/Index.aspx?DataSetCode=TABLECODE8021>. Accessed: 25 May 2017.

- [Stenberg, 2013] Stenberg, A. (2013). Interpreting estimates of heritability—a note on the twin decomposition. *Economics & Human Biology*, 11(2):201–205.
- [Stommel and Schoenborn, 2010] Stommel, M. and Schoenborn, C. A. (2010). Variations in bmi and prevalence of health risks in diverse racial and ethnic populations. *Obesity*, 18(9):1821–1826.
- [Sturm and An, 2014] Sturm, R. and An, R. (2014). Obesity and economic environments. *CA: A Cancer Journal for Clinicians*, 64(5):337–350.
- [Sturm and Datar, 2011] Sturm, R. and Datar, A. (2011). Regional price differences and food consumption frequency among elementary school children. *Public Health*, 125(3):136–141.
- [Swinburn et al., 2014] Swinburn, B., Dominick, C., and Vandevijvere, S. (2014). *Benchmarking food environments: experts' assessments of policy gaps and priorities for the New Zealand Government*. The University of Auckland, Auckland.
- [Tian et al., 2014] Tian, X., Zhao, G., Li, Y., Wang, L., and Shi, Y. (2014). Overweight and obesity difference of Chinese population between different urbanization levels: overweight and obesity between urbanization levels. *The Journal of Rural Health*, 30(1):101–112.
- [Vandevijvere et al., 2015] Vandevijvere, S., Chow, C. C., Hall, K. D., Umali, E., and Swinburn, B. A. (2015). Increased food energy supply as a major driver of the obesity epidemic: a global analysis. *Bulletin of The World Organization*, 93(7):446–456.
- [Vang et al., 2015] Vang, Z., Sigouin, J., Flenon, A., and Gagnon, A. (2015). The healthy immigrant effect in canada: a systematic review. *Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series*, 3(1):4.
- [Vartanian et al., 2007] Vartanian, L. R., Schwartz, M. B., and Brownell, K. D. (2007). Effects of soft drink consumption on nutrition and health: a systematic review and meta-analysis. *American Journal of Public Health*, 97(4):667–675.
- [Vazquez et al., 2007] Vazquez, G., Duval, S., Jacobs, D. R., and Silventoinen, K. (2007). Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiologic reviews*, 29(1):115–128.
- [Visscher et al., 2012] Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, 90(1):7–24.

- [Wadden et al., 2014] Wadden, T. A., Bantle, J. P., Blackburn, G. L., Bolin, P., Brancati, F. L., Bray, G. A., Clark, J. M., Coday, M., Dutton, G. R., Egan, C., Evans, M., Foreyt, J. P., Sengardi, S. G., Gregg, E. W., Hazuda, H. P., Hill, J. O., Horton, E. S., Hubbard, V. S., Jakicic, J. M., Jeffery, R. W., Johnson, K. C., Kahn, S. E., Kitabchi, A. E., Knowler, W. C., Lewis, C. E., Maschak-Carey, B. J., Montez, M. G., Montgomery, B., Nathan, D. M., Nelson, J., Patricio, J., Peters, A., Pi-Sunyer, F. X., Pownall, H., Rickman, A. D., Vitolins, M., Walkup, M. P., West, D. S., Williamson, D., Wing, R. R., Wyatt, H., Yanovski, S. Z., and Look Ahead Research Group (2014). Eight-year weight losses with an intensive lifestyle intervention: The look AHEAD study. *Obesity*, 22(1):5–13.
- [Wang et al., 2010] Wang, J., Williams, M., Rush, E., Crook, N., Forouhi, N. G., and Simmons, D. (2010). Mapping the availability and accessibility of healthy food in rural and urban New Zealand — Te Wai o Rona: Diabetes Prevention Strategy. *Public Health Nutrition*, 13(7):1049–1055.
- [Weiss and Kaufman, 2008] Weiss, R. and Kaufman, F. R. (2008). Metabolic complications of childhood obesity: identifying and mitigating the risk. *Diabetes care*, 31(2):S310–316.
- [Wing et al., 2010] Wing, R. R., Bahnson, J. L., Bray, G. A., Clark, J. M., Coday, M., Egan, C., Espeland, M. A., Foreyt, J. P., Gregg, E. W., Goldman, V., Haffner, S. M., Hazuda, H., Hill, J. O., Horton, E. S., Hubbard, V. S., Jakicic, J., Jeffery, R. W., Johnson, K. C., Kahn, S., Killean, T., Kitabchi, A. E., Lewis, C. E., Manus, C., Maschak-Carey, B. J., Michaels, S., Montez, M., Montgomery, B., Nathan, D. M., Patricio, J., Peters, A., Pi-Sunyer, X., Pownall, H., Reboussin, D., Rejeski, W. J., Rubin, R., Safford, M., Skarphol, T., Van Dorsten, B., Wadden, T. A., Wagenknecht, L., Wesche-Thobaben, J., West, D. S., Williamson, D., Yanovski, S. Z., and Group, L. A. R. (2010). Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the look AHEAD trial. *Archives of Internal Medicine*, 170(17):1566–1575.
- [Woelfel and Nestle, 2008] Woelfel, S. H. and Nestle, M. (2008). Do dietary guidelines explain the obesity epidemic? *American Journal of Preventive Medicine*, 34(3):263–265.
- [World Health Organization, 2015] World Health Organization (2015). Obesity and overweight. <http://www.who.int/mediacentre/factsheets/fs311/en/>. Accessed: 22 April 2017.
- [World Obesity Federation,] World Obesity Federation. Extended international (iotf) body mass index cut-offs for thinness, overweight and obesity in children. <http://www.worldobesity.org/data/cut-points-used/newchildcutoffs/>. Accessed: 13 July 2017.
- [Wu et al., 2015] Wu, S. Q., Ding, Y. Y., Wu, F. Q., Li, R. S., Hu, Y., Hou, J., and Mao, P. Y. (2015). Socio-economic position as an intervention against overweight and obesity in children: a systematic review and meta-analysis. *Scientific Reports*, 5:11354.

- [Xue et al., 2016] Xue, H., Wu, Y., Wang, X., and Wang, Y. (2016). Time trends in fast food consumption and its association with obesity among children in China. *PLoS ONE*, 11(3):e0151141.
- [Zhang et al., 2017] Zhang, J., Wang, D., Eldridge, A. L., Huang, F., Ouyang, Y., Wang, H., and Zhang, B. (2017). Urban-rural disparities in energy intake and contribution of fat and animal source foods in Chinese children aged 4-17 years. *Nutrients*, 9(5):526.

A R Codes

Listing A.1: Cleaning The NZHSs Data

```
1
2 ##### DATA CLEANING – New Zealand Health Surveys
3
4 ## Load NZHS data from 2002/03 to 2014/15 into the system. Subsetting is done in the Microsoft Excel
   2016.
5 # Subsetting done in the Microsoft Excel is not essential. It only helps to reduce to the time to
   load the data.
6 HS02 <- read.csv("D:/UC/Thesis/Data/NZHS/HS02-subset.csv", na.strings=c("", "NA"))
7 HS06A <- read.csv("D:/UC/Thesis/Data/NZHS/HS06A-subset.csv", na.strings=c("", "NA"))
8 HS06C <- read.csv("D:/UC/Thesis/Data/NZHS/HS06C-subset.csv", na.strings=c("", "NA"))
9 HS11A <- read.csv("D:/UC/Thesis/Data/NZHS/HS11A-subset.csv", na.strings=c("", "NA"))
10 HS11C <- read.csv("D:/UC/Thesis/Data/NZHS/HS11C-subset.csv", na.strings=c("", "NA"))
11 HS12A <- read.csv("D:/UC/Thesis/Data/NZHS/HS12A-subset.csv", na.strings=c("", "NA"))
12 HS12C <- read.csv("D:/UC/Thesis/Data/NZHS/HS12C-subset.csv", na.strings=c("", "NA"))
13 HS13A <- read.csv("D:/UC/Thesis/Data/NZHS/HS13A-subset.csv", na.strings=c("", "NA"))
14 HS13C <- read.csv("D:/UC/Thesis/Data/NZHS/HS13C-subset.csv", na.strings=c("", "NA"))
15 HS14A <- read.csv("D:/UC/Thesis/Data/NZHS/HS14A-subset.csv", na.strings=c("", "NA"))
16 HS14C <- read.csv("D:/UC/Thesis/Data/NZHS/HS14C-subset.csv", na.strings=c("", "NA"))
17
18
19 ##### Install Packages #####
20
21 # package to do crosstable
22 install.packages("gmodels")
23
24 #package to draw chart, graph, etc.
25 install.packages("ggplot2")
26
27 # package for text dodge in charts and graphs
28 install.packages("ggrepel")
29
30 # package for a colour pallette
31 install.packages("RColorBrewer")
32
```

```
33 # package for creating mosaic plot
34 install.packages("ggmosaic")
35
36 # package for transforming and re-arranging data
37 install.packages("dplyr")
38 install.packages("tidyr")
39
40 # package for wrapping long text
41 install.packages("stringr")
42
43 # package to arrange plot
44 install.packages("gridExtra")
45
46 # package to adjust plots' size in gridExtra
47 install.packages('cowplot')
48
49 # function for multiple crosstabs
50 source("http://pcwww.liv.ac.uk/~william/R/crosstab.r")
51
52 # package for confidence interfal and effect size calculator
53 install.packages("MBESS")
54
55 # package for creating a survey desing object
56 install.packages("survey")
57 install.packages("segmented")
58
59 # package for reading a sas7bdat format
60 install.packages("sas7bdat")
61
62 # package for cutting strings
63 install.packages("stringi")
64
65 # package for calculating kurtosis and skewness
66 install.packages("moments")
67
68 # package for evaluating environment inside tidyr and dyplr functions
69 install.packages("lazyeval")
70
71 # package for drawing interactive charts
72 install.packages("plotly")
```

```

73
74
75 ##### Load the packages #####
76 library(gmodels)
77 library(ggplot2)
78 library(ggmosaic)
79 library(ggrepel)
80 library(RColorBrewer)
81 library(stringi)
82 library(coin)
83 library(scales)
84 library(moments)
85 library(tidyr)
86 library(plyr)
87 library(dplyr)
88 library(stringr)
89 library(gridExtra)
90 library(cowplot)
91 library(lazyeval)
92 library(MBESS)
93 library(plotly)
94 library(survey)
95 library(segmented)
96
97
98 ##### Function Sets #####
99 ## Function set 1 (proportion table and several statistic tests)
100
101 # Colourblind Paleta
102 GreyPalette <- c("#d3d3d3", "#bdbdbd", "#a8a8a8", "#939393", "#7e7e7e", "#696969", "#545454", "#3f3f3f", "#2
    a2a2a", "#151515", "#000000")
103 cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#
    CC79A7")
104
105
106 # Cramer's v function
107 cv.test <- function(x,y) {
108   # Cramer's V function
109   CV = sqrt(chisq.test(x, y, correct=FALSE)$statistic /
110     (length(x) * (min(length(unique(x)), length(unique(y))) - 1)))

```

```

111   return(as.numeric(CV))
112 }
113
114
115 # A function for non-linear test
116 non.lin <- function(x,y){
117   # non-linear association chi square
118   X2 = (chisq.test(x, y, correct=FALSE)$statistic) - (statistic(lbl_test(table(x,y)))^2)
119   p = 1-pchisq(as.numeric(X2),(min(length(unique(x)),length(unique(y))) - 1))
120   print.noquote("X2 non linear & p value")
121   return(cbind(as.numeric(X2),as.numeric(p)))
122 }
123
124
125 # A function for creating proportion table by groups
126 prop_frame <- function(data, var1) {
127   # a function to show number and proportion of grouped variables , use quote on var1
128
129   df <- data %>%
130     group_by_(var1) %>%
131     filter_(interp(~ !is.na(var), var = as.name(var1))) %>%
132     summarise(n=n()) %>%
133     mutate(prop=n/sum(n))
134
135   df$perc = round(df$prop*100, 2)
136   df$cop = sprintf("%s (%s)", df$n, df$perc)
137
138   return(df)
139 }
140
141
142
143 ## Function set 2 (normal plot, charts, graphs)
144
145 # A function for displaying histogram and qq plot
146 normal_plot <- function(data, var1){
147   # histogram and qq plot function. Use quote for var1 (eg. var1 = "bmiscale")
148
149   vector <- as.numeric(unlist(data[var1]))
150

```



```

151 histogram <- ggplot(data, aes_string(var1)) +
152   geom_histogram(aes(y=..density..), bins= 50, alpha= 0.5, na.rm= TRUE) +
153   geom_density(col=2, na.rm= TRUE) +
154   scale_x_continuous(limits=c(18,95), breaks=seq(10,100,10)) +
155   scale_y_continuous()
156
157
158 y <- quantile(vector[!is.na(vector)], c(0.25, 0.75))
159 x <- qnorm(c(0.25, 0.75))
160 slope <- diff(y)/diff(x)
161 int <- y[1L] - (slope*x[1L])
162
163 qq <- ggplot(data, aes_string(sample=var1)) +
164   stat_qq() +
165   geom_abline(slope= slope, intercept= int, color="red") +
166   scale_y_continuous()
167
168 grid <- grid.arrange(histogram, qq, ncol=2)
169
170 return(grid)
171 }
172
173
174
175 # function for box plot + showing median
176 box_plot_median <- function(data, var_x, var_y, title_x, title_y, year, age=">=18 yo", txt.size = 3,
177   t.size = 15, ax.size = 10, x.size = 10, y.size = 10,
178   min=0, max=100, bins = 10, wide=0.8, gap= 2){
179   # box plot for numerical scale by categorical group, use quote for var_x ~ year
180   # year is just a string of text for the title
181
182   complete_data <- data %>%
183     select_(var_x, var_y) %>%
184     filter_(interp(~!is.na(var), var= as.name(var_x)))
185
186   colnames(complete_data) <- c("var1", "var2")
187
188   frame <- complete_data %>%
189     group_by(var1) %>%
190     summarise_each(funs(median), var2)

```

```

191
192 colnames(frame)[2] <- "number"
193
194 plot <- ggplot(complete_data, aes(x=var1, y=var2, na.rm=TRUE), environment=environment()) +
195   geom_boxplot(width=wide) +
196   scale_y_continuous(breaks=seq(0,100,bins), limits=c(min,max)) +
197   labs(title=sprintf("%s by %s \n(%s, age)", title_y, title_x, year),
198        x= sprintf("%s", title_x), y= sprintf("%s", title_y)) +
199   theme(axis.title = element_text(size= ax.size, face="bold"),
200         plot.title= element_text(size= t.size, face="bold", hjust=0.5),
201         axis.text.x = element_text(size = x.size),
202         axis.text.y = element_text(size = y.size)) +
203   geom_text(data=frame, aes(label=round(frame$number,1)), y=frame$number+gap, size= txt.size) +
204   scale_x_discrete(label= function(x) str_wrap(x, width=10))
205
206 return(plot)
207 }
208
209
210 # box chart without title
211 box_notitle <- function(data, var_x, var_y, title_x, title_y, txt.size = 3,
212                        t.size = 10, ax.size = 6, x.size = 8, y.size = 8,
213                        min=0, max=100, bins = 10, wide=0.8, gap= 3){
214   # box plot for numerical scale by categorical group, use quote for var_x ~ year
215   # year is just a string of text for the title
216
217   complete_data <- data %>%
218     select_(var_x, var_y) %>%
219     filter_(interp(~!is.na(var), var= as.name(var_x)))
220
221   colnames(complete_data) <- c("var1", "var2")
222
223   frame <- complete_data %>%
224     group_by(var1) %>%
225     summarise_each(funs(median), var2)
226
227   colnames(frame)[2] <- "number"
228
229   plot <- ggplot(complete_data, aes(x=var1, y=var2, na.rm=TRUE), environment=environment()) +
230     geom_boxplot(width=wide) +

```

```

231 scale_y_continuous(breaks=seq(0,100,bins), limits=c(min,max)) +
232 labs(title=sprintf("%s by %s", title_y, title_x),
233       x= sprintf("%s", title_x), y= sprintf("%s", title_y)) +
234 theme(axis.title = element_text(size= ax.size, face="bold"),
235        plot.title= element_text(size= t.size, face="bold", hjust=0.5),
236        axis.text.x = element_text(size = x.size),
237        axis.text.y = element_text(size = y.size)) +
238 geom_text(data=frame, aes(label=round(frame$number,1)), y=frame$number+gap, size= txt.size) +
239 scale_x_discrete(label= function(x) str_wrap(x, width=10))
240
241 return(plot)
242 }
243
244
245 # function for stacked plot + showing percentage
246 stacked_plot <- function(data, var_x, var_y, title1, title2, year, age = ">=18 yo", ax.size = 10,
247                           t.size = 15, l.size = 10, x.size = 10, y.size = 10, txt.size = 3){
248   # stacked bar chart, use quote for var_x ~ age
249   # year is just a string of text
250
251   data_frame <- data %>%
252     filter_(interp(~ !is.na(var), var = as.name(var_x)),
253              interp(~ !is.na(var), var = as.name(var_y))) %>%
254     group_by_(var_x, var_y) %>%
255     summarise(count=n()) %>%
256     mutate(percent = count/sum(count),
257            per = count/sum(count)*100,
258            pos = (cumsum(percent) - 0.5*percent)) # calculating position
259
260   # labels
261   data_frame$label_percent <- paste0(sprintf("%.1f", data_frame$per), "%")
262
263   colnames(data_frame)[1:2] <- c("var_x", "var_y")
264
265   localenv <- environment()
266
267   plot <- ggplot(data_frame, aes(x= var_x, y=percent, fill=factor(var_y, levels=rev(var_y))),
268                  environment= localenv) +
269   geom_bar(stat="identity", width=0.8) +
270   labs(title= sprintf("%s by %s \n(%, %s)", title1, title2, year, age),

```

```

270     x= sprintf("%s", title2), y= "Density") +
271     scale_fill_discrete(name = sprintf("%s", title1)) +
272     theme(axis.title = element_text(size = ax.size, face="bold"),
273           plot.title = element_text(size = t.size, face="bold", hjust=0.5),
274           legend.title = element_text(size = l.size, face="bold"),
275           axis.text.x = element_text(size = x.size),
276           axis.text.y = element_text(size = y.size)) +
277     geom_label(aes(y=data_frame$pos, label=data_frame$label_percent), size=t.txt.size, fill="White") +
278     scale_x_discrete(label= function(x) str_wrap(x, width=8))
279
280     return(plot)
281 }
282
283 # stacked bar chart without title
284 stacked_notitle <- function(data, var_x, var_y, title1, title2, ax.size = 6,
285                             t.size = 10, l.size = 8, x.size = 8, y.size = 8, lt.size = 8, txt.size = 3)
286     {
287     # stacked bar chart, use quote for var_x ~ age
288     # year is just a string of text
289
290     data_frame <- data %>%
291       filter_(interp(~ !is.na(var), var = as.name(var_x)),
292               interp(~ !is.na(var), var = as.name(var_y))) %>%
293       group_by_(var_x, var_y) %>%
294       summarise(count=n()) %>%
295       mutate(percent = count/sum(count),
296              per = count/sum(count)*100,
297              pos = (cumsum(percent) - 0.5*percent)) # calculating position
298
299     # labels
300     data_frame$label_percent <- paste0(sprintf("%.1f", data_frame$per), "%")
301
302     colnames(data_frame)[1:2] <- c("var_x", "var_y")
303
304     localenv <- environment()
305
306     plot <- ggplot(data_frame, aes(x= var_x, y=percent, fill=factor(var_y, levels=rev(var_y))),
307                   environment= localenv) +
308     geom_bar(stat="identity", width=0.8) +
309     labs(title= sprintf("%s by %s", title1, title2),

```

```

308     x= sprintf("%s", title2), y= "Density") +
309     scale_fill_discrete(name = sprintf("%s", title1)) +
310     theme(axis.title = element_text(size = ax.size, face="bold"),
311           plot.title = element_text(size = t.size, face="bold", hjust=0.5),
312           legend.title = element_text(size = l.size, face="bold"),
313           axis.text.x = element_text(size = x.size),
314           axis.text.y = element_text(size = y.size),
315           legend.text=element_text(size = lt.size)) +
316     geom_text(aes(y=data_frame$pos, label=data_frame$label_percent), size=t.txt.size) +
317     scale_x_discrete(label= function(x) str_wrap(x, width=8))
318
319   return(plot)
320 }
321
322
323 # stacked bar chart for ethnicity
324 stacked_eth <- function(data, var_x, var_y, title1, title2, ax.size = 6,
325                          t.size = 10, l.size = 8, x.size = 8, y.size = 8, lt.size = 8, txt.size = 2)
326   {
327     # stacked bar chart, use quote for var_x ~ age
328     # year is just a string of text
329
330     data_frame <- data %>%
331       filter_(interp(~ !is.na(var), var = as.name(var_x)),
332               interp(~ !is.na(var), var = as.name(var_y))) %>%
333       group_by_(var_x, var_y) %>%
334       summarise(count=n()) %>%
335       mutate(percent = count/sum(count),
336              per = count/sum(count)*100,
337              pos = (cumsum(percent) - 0.5*percent)) # calculating position
338
339     # labels
340     data_frame$label_percent <- paste0(sprintf("%.1f", data_frame$per), "%")
341
342     colnames(data_frame)[1:2] <- c("var_x", "var_y")
343
344     localenv <- environment()
345
346     plot <- ggplot(data_frame, aes(x= var_x, y=percent, fill=factor(var_y, levels=rev(var_y))),
347                    environment= localenv) +

```

```

346 geom_bar(stat="identity", width=0.8) +
347 labs(title= sprintf("%s by %s", title1, title2),
348       x= sprintf("%s", title2), y= "Density") +
349 scale_fill_discrete(name = sprintf("%s", title1)) +
350 theme(axis.title = element_text(size = ax.size, face="bold"),
351        plot.title = element_text(size = t.size, face="bold", hjust=0.5),
352        legend.title = element_text(size = l.size, face="bold"),
353        axis.text.x = element_text(size = x.size),
354        axis.text.y = element_text(size = y.size),
355        legend.text=element_text(size = lt.size)) +
356 geom_text(aes(y=data_frame$pos, label=data_frame$label_percent), size=txt.size, position =
357           position_dodge(width=1)) +
358
359 scale_x_discrete(label= function(x) str_wrap(x, width=8))
360
361 return(plot)
362 }
363
364 # stacked bar chart without label
365 stacked_nolabel <- function(data, var_x, var_y, title1, title2, year){
366   # stacked bar chart without label, use quote for var_x ~ year
367   # year is just a string of text
368
369   data_frame <- data %>%
370     filter_(interp(~ !is.na(var), var = as.name(var_x)), interp(~ !is.na(var), var = as.name(var_y))
371              ) %>%
372     group_by_(var_x, var_y) %>%
373     summarise(count=n()) %>%
374     mutate(percent = count/sum(count),
375            per = count/sum(count)*100)
376
377   colnames(data_frame)[1:2] <- c("var_x", "var_y")
378
379   localenv <- environment()
380
381   plot <- ggplot(data_frame, aes(x= var_x, y=percent, fill=var_y), environment= localenv) +
382     geom_bar(stat="identity", width=0.8) +
383     labs(title= sprintf("%s by %s \n(%s, >=18 yo)", title2, title1, year),
384          x= sprintf("%s", title1), y= "Density") +

```

```

384   scale_fill_discrete(name = sprintf("%s", title2)) +
385   theme(axis.title = element_text(face="bold"), plot.title= element_text(face="bold", hjust=0.5),
386         legend.title=element_text(face="bold")) +
387   scale_x_discrete(label= function(x) str_wrap(x, width=10))
388
389   return(plot)
390 }
391
392
393
394 ## Function set 3 (custom statistic tables + effect sizes)
395
396 # function for anova table + effect size confidence interval
397 anova_table <- function(df1, df2){
398   # neat anova table with effect size CI
399   df_between <- anova(aov(df1 ~ df2))[1,1]
400   df_resid <- anova(aov(df1 ~ df2))[2,1]
401   F_value <- anova(aov(df1 ~ df2))[1,4]
402   p_value <- anova(aov(df1 ~ df2))[1,5]
403   eta_squared <- anova(aov(df1 ~ df2))[1,2]/sum(anova(aov(df1 ~ df2))[ ,2])
404
405   es_ci <- as.data.frame(ci.pvaf(F.value= F_value, df.1= df_between,
406                                 df.2= df_resid, N= df_between+df_resid+1, conf.level= .90))
407
408   lower_ci <- es_ci[1,1]
409   upper_ci <- es_ci[1,3]
410
411   table <- cbind(F_value, df_between, p_value, eta_squared, lower_ci, upper_ci)
412
413   return(table)
414 }
415
416
417 # function for chi square table + effect size confidence interval
418 chix_table <- function(df1, df2){
419   # nice chi squared table with cramer's V and CI
420   X2 <- chisq.test(df1, df2, correct=FALSE)[1]
421   df <- chisq.test(df1, df2, correct=FALSE)[2]
422   p_value <- chisq.test(df1, df2, correct=FALSE)[3]
423

```

```

424 # Craner's V formula in another function (= cv.test)
425 cramer_v <- cv.test(df1, df2)
426
427 # transform Cramer's V into Fisher's Z
428 fisher_z <- 0.5 * log((1 + cramer_v)/(1 - cramer_v))
429
430 # calculate sample size
431 N_size <- sum(table(df1, df2))
432
433 # calculate standard error per Fisher's Z distribution
434 standard_error <- 1/sqrt(sum(N_size)-3) * qnorm(1-(0.05/2))
435
436 # confidence interval around Fisher's Z
437 ci_fz <- fisher_z + c(-standard_error, standard_error)
438
439 # confidence interval around Cramer's V
440 ci_crv <- (exp(2 * ci_fz) - 1)/(1 + exp(2 * ci_fz))
441
442 # bind them all in a table
443 table <- cbind(X2, df, p_value, cramer_v, ci_crv[1], ci_crv[2])
444
445 colnames(table)[5] <- "Lower V"
446 colnames(table)[6] <- "Upper V"
447
448 table[,c(1,4:6)] <- lapply(table[,c(1,4:6)], function(x){round(x,4)})
449
450 return(table)
451 }
452
453
454 # function for wilcoxon test table
455 wilcox_table <- function(df1, df2){
456   # Wilcoxon rank-sum test with ties summary in a nice table
457
458   W <- wilcox.test(df1 ~ df2, correct=TRUE)$statistic
459   p_value <- wilcox.test(df1 ~ df2, correct = TRUE)$p.value
460   # which method should I use, idk!?!?!?
461
462   cohen_d <- qnorm(p_value)/sqrt(length(df1)) # effect size using Rosenthal's Formula
463

```



```

464 table <- cbind(W, p_value , cohen_d)
465
466 return(table)
467
468 }
469
470
471 # A function to display OR, 95% CL, z score, and p value from the regression summary output (svyolr)
472 summy <- function(x, round=3) {
473   # A function to display OR, CL, and t value from 'svyolr'
474   beta <- matrix(coefficients(summary(x))[,1])
475
476   z <- matrix(coefficients(summary(x))[,3])
477
478   p <- matrix(2*pnorm(-abs(z))) # two-tailed z-test
479
480   matrix <- cbind(matrix(beta), matrix(exp(coef(x))), exp(confint(x)), z, p)
481
482   dimnames(matrix)[[2]][1] <- "beta"
483   dimnames(matrix)[[2]][2] <- "OR"
484   dimnames(matrix)[[2]][5] <- "z-score"
485   dimnames(matrix)[[2]][6] <- "p-value"
486
487   return(round(matrix,round))
488 }
489
490
491 ## Function set 4 (custom crosstable)
492
493 # Crosstable functions
494 xt_2x3 <- function(df1, df2){
495   # 3x3 table
496   main_table <- table(df1, df2)
497
498   r_percent <- round(prop.table(main_table,1),4)
499
500   r1_total <- sum(main_table[1,])
501   r2_total <- sum(main_table[2,])
502
503   row_total <- r1_total + r2_total

```

```

504
505 # row percentage (4th column)
506 per_r1 <- round((r1_total/row_total),4)
507 per_r2 <- round((r2_total/row_total),4)
508
509 # 4th column
510 col4_total <- c(r1_total , per_r1 , r2_total , per_r2)
511
512 # column total
513 c1_total <- sum(main_table[,1])
514 c2_total <- sum(main_table[,2])
515 c3_total <- sum(main_table[,3])
516
517 # total N
518 total_N <- sum(main_table)
519
520 row5_total <- c(c1_total , c2_total , c3_total , total_N)
521
522 # combine everything into one table
523 xtab <- rbind(as.numeric(main_table[1,]), r_percent[1,], as.numeric(main_table[2,]), r_percent
524               [2,])
525
526 xtab1 <- cbind(xtab , col4_total)
527
528 xtab2 <- rbind(xtab1 , row5_total)
529
530 # rename row and col names
531 row_names <- rownames(main_table)
532
533 rownames(xtab2) <- c(row_names[1], "%", row_names[2], "%", "Column Total")
534 colnames(xtab2)[4] <- "Row Total"
535
536 return(xtab2)
537 }
538
539 xt_3x3 <- function(df1 , df2){
540 # 3x3 table
541 main_table <- table(df1 , df2)
542

```

```

543 r_percent <- round(prop.table(main_table,1),4)
544
545 r1_total <- sum(main_table[1,])
546 r2_total <- sum(main_table[2,])
547 r3_total <- sum(main_table[3,])
548
549 row_total <- r1_total + r2_total + r3_total
550
551 # row percentage (4th column)
552 per_r1 <- round((r1_total/row_total),4)
553 per_r2 <- round((r2_total/row_total),4)
554 per_r3 <- round((r3_total/row_total),4)
555
556 # 4th column
557 col4_total <- c(r1_total, per_r1, r2_total, per_r2, r3_total, per_r3)
558
559 # column total
560 c1_total <- sum(main_table[,1])
561 c2_total <- sum(main_table[,2])
562 c3_total <- sum(main_table[,3])
563
564 # total N
565 total_N <- sum(main_table)
566
567 row5_total <- c(c1_total, c2_total, c3_total, total_N)
568
569 # combine everything into one table
570 xtab <- rbind(as.numeric(main_table[1,]), r_percent[1,], as.numeric(main_table[2,]), r_percent
571               [2,],
572               as.numeric(main_table[3,]), r_percent[3,])
573
574 xtab1 <- cbind(xtab, col4_total)
575
576 xtab2 <- rbind(xtab1, row5_total)
577
578 # rename row and col names
579 row_names <- rownames(main_table)
580
581 rownames(xtab2) <- c(row_names[1], "%", row_names[2], "%", row_names[3], "%", "Column Total")
582 colnames(xtab2)[4] <- "Row Total"

```

```

582
583   return(xtab2)
584
585 }
586
587 xt_4x3 <- function(df1, df2){
588   # 4x3 table
589   main_table <- table(df1, df2)
590
591   r_percent <- round(prop.table(main_table,1),4)
592
593   r1_total <- sum(main_table[1,])
594   r2_total <- sum(main_table[2,])
595   r3_total <- sum(main_table[3,])
596   r4_total <- sum(main_table[4,])
597
598   row_total <- r1_total + r2_total + r3_total + r4_total
599
600   # row percentage (4th column)
601   per_r1 <- round((r1_total/row_total),4)
602   per_r2 <- round((r2_total/row_total),4)
603   per_r3 <- round((r3_total/row_total),4)
604   per_r4 <- round((r4_total/row_total),4)
605
606   # 4th column
607   col4_total <- c(r1_total, per_r1, r2_total, per_r2, r3_total, per_r3, r4_total, per_r4)
608
609   # column total
610   c1_total <- sum(main_table[,1])
611   c2_total <- sum(main_table[,2])
612   c3_total <- sum(main_table[,3])
613
614   # total N
615   total_N <- sum(main_table)
616
617   row5_total <- c(c1_total, c2_total, c3_total, total_N)
618
619   # combine everything into one table
620   xtab <- rbind(as.numeric(main_table[1,]), r_percent[1,], as.numeric(main_table[2,]), r_percent
    [2,],

```

```

621         as.numeric(main_table[3,]), r_percent[3,], as.numeric(main_table[4,]), r_percent
622         [4,])
623
624     xtab1 <- cbind(xtab, col4_total)
625
626     xtab2 <- rbind(xtab1, row5_total)
627
628     # rename row and col names
629     row_names <- rownames(main_table)
630
631     rownames(xtab2) <- c(row_names[1], "%", row_names[2], "%", row_names[3], "%", row_names[4], "%", "
        Column Total")
632
633     colnames(xtab2)[4] <- "Row Total"
634
635     return(xtab2)
636 }
637
638 xt_5x3 <- function(df1, df2){
639     # 5x3 table
640     main_table <- table(df1, df2)
641
642     r_percent <- round(prop.table(main_table, 1), 4)
643
644     r1_total <- sum(main_table[1,])
645     r2_total <- sum(main_table[2,])
646     r3_total <- sum(main_table[3,])
647     r4_total <- sum(main_table[4,])
648     r5_total <- sum(main_table[5,])
649
650     row_total <- r1_total + r2_total + r3_total + r4_total + r5_total
651
652     # row percentage (4th column)
653     per_r1 <- round((r1_total/row_total), 4)
654     per_r2 <- round((r2_total/row_total), 4)
655     per_r3 <- round((r3_total/row_total), 4)
656     per_r4 <- round((r4_total/row_total), 4)
657     per_r5 <- round((r5_total/row_total), 4)
658
659     # 4th column

```

```

659 col4_total <- c(r1_total , per_r1 , r2_total , per_r2 , r3_total , per_r3 , r4_total , per_r4 , r5_total ,
660               per_r5)
661
662 # column total
663 c1_total <- sum(main_table[,1])
664 c2_total <- sum(main_table[,2])
665 c3_total <- sum(main_table[,3])
666
667 # total N
668 total_N <- sum(main_table)
669
670 row5_total <- c(c1_total , c2_total , c3_total , total_N)
671
672 # combine everything into one table
673 xtab <- rbind(as.numeric(main_table[1,]) , r_percent[1,] , as.numeric(main_table[2,]) , r_percent
674               [2,] ,
675               as.numeric(main_table[3,]) , r_percent[3,] , as.numeric(main_table[4,]) , r_percent
676               [4,] ,
677               as.numeric(main_table[5,]) , r_percent[5,])
678
679 xtab1 <- cbind(xtab , col4_total)
680
681 xtab2 <- rbind(xtab1 , row5_total)
682
683 # rename row and col names
684 row_names <- rownames(main_table)
685
686 rownames(xtab2) <- c(row_names[1], "%", row_names[2], "%", row_names[3], "%", row_names[4], "%",
687                     row_names[5], "%", "Column Total")
688 colnames(xtab2)[4] <- "Row Total"
689
690 return(xtab2)
691
692 }
693
694 ##### NZHS 2002/03 – 15 years and over #####

```

```

695 ## remove clothing adjustment, added 1.2 kg
696 HS02$Weight <- HS02$Weight + 1.2
697
698 ## recalculate BMI
699 HS02$bmi <- HS02$Weight/((HS02$Height/100)^2)
700
701 ## Create a new bmi category using IOTF cut off
702 # 0 = underweight; 1= average; 2= overweight; 3=obese
703 HS02 <- HS02 %>%
704   mutate(bmic = ifelse(Sex == "M" & Age == 15 & bmi <16.98, 0,
705     ifelse(Sex == "M" & Age == 16 & bmi <17.53, 0,
706       ifelse(Sex == "M" & Age == 17 & bmi <18.04, 0,
707         ifelse(Sex == "F" & Age == 15 & bmi <17.43, 0,
708           ifelse(Sex == "F" & Age == 16 & bmi <17.9, 0,
709             ifelse(Sex == "F" & Age == 17 & bmi <18.24, 0,
710               ifelse(Age >= 18 & bmi <18.5, 0,
711                 ifelse(Sex == "M" & Age == 15 & bmi >= 16.98 & bmi <23.28, 1,
712                   ifelse(Sex == "M" & Age == 16 & bmi >= 17.53 & bmi <23.89, 1,
713                     ifelse(Sex == "M" & Age == 17 & bmi >= 18.04 & bmi <24.46, 1,
714                       ifelse(Sex == "F" & Age == 15 & bmi >= 17.43 & bmi <23.89, 1,
715                         ifelse(Sex == "F" & Age == 16 & bmi >= 17.90 & bmi <24.34, 1,
716                           ifelse(Sex == "F" & Age == 17 & bmi >= 18.24 & bmi <24.70, 1,
717                             ifelse(Age >= 18 & bmi >= 18.5 & bmi <25, 1,
718                               ifelse(Sex == "M" & Age == 15 & bmi >= 23.28 & bmi <28.32, 2,
719                                 ifelse(Sex == "M" & Age == 16 & bmi >= 23.89 & bmi <28.89, 2,
720                                   ifelse(Sex == "M" & Age == 17 & bmi >= 24.46 & bmi <29.43, 2,
721                                     ifelse(Sex == "F" & Age == 15 & bmi >= 23.89 & bmi <29.01, 2,
722                                       ifelse(Sex == "F" & Age == 16 & bmi >= 24.34 & bmi <29.40, 2,
723                                         ifelse(Sex == "F" & Age == 17 & bmi >= 24.70 & bmi <29.70, 2,
724                                           ifelse(Age >= 18 & bmi >= 25 & bmi <30, 2,
725                                             ifelse(Sex == "M" & Age == 15 & bmi >=28.32, 3,
726                                               ifelse(Sex == "M" & Age == 16 & bmi >=28.89, 3,
727                                                 ifelse(Sex == "M" & Age == 17 & bmi >=29.43, 3,
728                                                   ifelse(Sex == "F" & Age == 15 & bmi >=29.01, 3,
729                                                     ifelse(Sex == "F" & Age == 16 & bmi >=29.40, 3,
730                                                       ifelse(Sex == "F" & Age == 17 & bmi >=29.70, 3,
731                                                         ifelse(Age >= 18 & bmi >= 30, 3 ,NA))))))))))))))))))))))))))
732
733
734 HS02$bmic <- as.factor(HS02$bmic)

```

```

735 summary(HS02$bmic) # 186 underweight
736
737
738 ## Exclude underweight
739 hs02 <- subset(HS02, !bmic == 0 | is.na(bmic))
740
741
742 ## Sex
743 hs02$gender <- factor(hs02$Sex, labels=c("Female", "Male"))
744 hs02$gender <- factor(hs02$gender, levels=c("Male", "Female"))
745
746 prop_frame(hs02, "gender")
747
748
749 ## Age
750 summary(hs02$Age)
751
752 # change the variable name to match the rest of NZHS data
753 names(hs02)[names(hs02) == "Age"] <- "age"
754
755
756 ## Migration Status
757 # Migrant = less than 10–11 years of residing in NZ
758 hs02$Q266 <- as.factor(hs02$Q266)
759
760 hs02$native <- recode(hs02$Q266, "1993"="Migrant", " 1998" = "Migrant", "1999"="Migrant",
761                             " 2000"="Migrant", "2001"="Migrant", "2002"="Migrant",
762                             " 2003"="Migrant", "2004"="Migrant", "2005"="Migrant",
763                             "2006"="Migrant", "2007"="Migrant", .default="Native")
764
765 hs02$native[is.na(hs02$native)] <- "Native"
766
767 prop_frame(hs02, "native")
768
769
770
771 ## BMI
772 # rename to match the rest of NZHS
773 hs02$bmiscale <- as.numeric(as.character(hs02$bmi))
774

```



```

775 # checking for normality
776 normal_plot(hs02, "bmic")
777
778 # label them
779 hs02$bmic <- factor(hs02$bmic, labels=c("Average", "Overweight", "Obese"))
780
781 prop_frame(hs02, "bmic")
782
783
784 ## Ethnicity
785 # Recode anyone ~ European
786 hs02 = hs02 %>%
787   mutate(euro = ifelse(Q263_01 == 1, 1, 0))
788
789 # Recode anyone ~ Maori
790 hs02 = hs02 %>%
791   mutate(maori = ifelse(Q263_02 == 1, 1, 0))
792
793 # Recode anyone ~ Pacific
794 hs02 = hs02 %>%
795   mutate(pacific = ifelse(Q263_03 == 1, 1,
796     ifelse(Q263_04 == 1, 1,
797       ifelse(Q263_05 == 1, 1,
798         ifelse(Q263_06 == 1, 1,
799           ifelse(Q263_07 == 1, 1, 0))))))
800
801 # Recode anyone ~ asian (chinese and indian only)
802 hs02 = hs02 %>%
803   mutate(asian = ifelse(Q263_08 == 1, 1,
804     ifelse(Q263_09 == 1, 1, 0)))
805
806 # Recode anyone ~ other (koreana, sri langkan, other asian are here)
807 hs02 = hs02 %>%
808   mutate(other = ifelse(Q263_10 == 1, 1,
809     ifelse(Q263_11 == 1, 1,
810       ifelse(Q263_12 == 1, 1,
811         ifelse(Q263_13 == 1, 1,
812           ifelse(Q263_14 == 1, 1,
813             ifelse(Q263_Other == 1, 1, 0)))))))
814

```

```

815 # Recode them into: Maori only, Pacific only, Asian only*, European only, Other, 2+ ethnicities (M)
816 hs02 = hs02 %>%
817   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
      Only",
818     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
      Pacific Only",
819     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
      Only",
820     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
      European Only",
821     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
      ,
822     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
      Ethnicities (M)",
823     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
      Ethnicities (M)",
824     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
      ,
825     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
      ,
826     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
      ,
827     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
      Ethnicities (M)",
828     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
      Ethnicities (M)",
829     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
      ,
830     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
      ,
831     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
      ,
832     ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
      Ethnicities (M)",
833     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
      Ethnicities (M)",
834     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
      Ethnicities (M)",
835     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
      ,

```

```

836         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
837         ,
838         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
839         ,
840         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
            Ethnicities (M)",
841         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
            Ethnicities (M)",
842         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
            Ethnicities (M)",
843         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
844         ,
845         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
            Ethnicities (M)",
846         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
            Ethnicities (M)",
847         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
            Ethnicities (M)",
848         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
849         ,
850         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
            Ethnicities (M)",
851         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
            Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))))))
852
853 hs02$eth_count <- as.factor(hs02$eth_count)
854 prop_frame(hs02, "eth_count")
855
856 ## Household Income
857 hs02$Q293 <- as.factor(hs02$Q293)
858 hs02$Q293_imputeflag <- as.factor(hs02$Q293_imputeflag)
859
860 # remove imputation
861 hs02 <- hs02 %>%
862   mutate(hhinc = ifelse(Q293_imputeflag == 1, as.numeric(as.character(Q293)), NA))
863
864 hs02$hhinc <- as.factor(hs02$hhinc)
865 summary(hs02$hhinc) # made sure the code works

```

```

864 # missing data
865 sum(is.na(hs02$hhinc[hs02$age >=18]))
866 sum(is.na(hs02$hhinc[hs02$age >=18])/length(hs02$hhinc[hs02$age >=18]))
867
868 hs02$hhinc <- as.numeric(as.character(hs02$hhinc))
869 hs02$hhinc <- cut(hs02$hhinc, breaks=c(0,5,9,11,13))
870 hs02$hhinc <- factor(hs02$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000",
871                                           ">$70,000"))
872
873 prop_frame(hs02, "hhinc")
874
875
876 ## Education
877 # recode 11 = refused to answer into NA
878 hs02$Q279 <- as.factor(hs02$Q279)
879 hs02$Q279[hs02$Q279 == "11"] <- NA
880
881 # recode secondary qualification
882 hs02$secondary <- recode(hs02$Q279, "1" = "1", .default = "2") # 1= no degree, 2= secondary school
      grads
883
884
885 # recode 97 = refused to answer into NA
886 hs02$Q280 <- as.factor(hs02$Q280)
887 hs02$Q280[hs02$Q280 == "97"] <- NA # 97: refused to answer
888
889 # recode tertiary qualification
890 hs02$tertiary <- recode(hs02$Q280, "99" = "0", .default = "3") #99 = no tertiary, 3= tertiary
891 summary(hs02$tertiary)
892
893 # recode educational qualification
894 hs02 <- mutate(hs02, edu = ifelse(secondary == "1" & tertiary == "0", 1,
895                                   ifelse(secondary == "2" & tertiary == "0", 2,
896                                           ifelse(tertiary == "3", 3, 1))))
897
898 hs02$edu <- as.factor(hs02$edu)
899 hs02$edu <- factor(hs02$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
900
901 prop_frame(hs02, "edu")
902

```

```

903
904 ## Deprivation Quintile
905 # recode XX into NA
906 summary(hs02$Nzdep01_quintiles)
907 hs02$Nzdep01_quintiles[hs02$Nzdep01_quintiles == "XX"] <- NA
908
909 # rearrange
910 hs02$Nzdep01_quintiles <- factor(hs02$Nzdep01_quintiles, labels=c("1","2","3","4","5"))
911
912 prop_frame(hs02, "Nzdep01_quintiles")
913
914 # rename
915 names(hs02)[names(hs02) == "Nzdep01_quintiles"] <- "dep"
916
917
918 ## Urban/Rural area
919 # collapse them into two category
920 hs02$UA <- recode(hs02$UA_type1, "Rural Centre"="Rural", "Rural"="Rural", .default="Urban")
921
922 prop_frame(hs02, "UA")
923
924
925 ## Alcohol Problem
926 hs02$Audit <- as.numeric(as.character(hs02$Audit))
927
928 # recode them into alcohol problem status (yes or no)
929 hs02$haz_drinker_all <- cut(hs02$Audit, breaks= c(-Inf, 7, Inf), labels = c("No Alcohol Problem", "
    Alcohol Problem"))
930
931 prop_frame(hs02, "haz_drinker_all")
932
933
934 ## Physical Activity
935 # Q139: how many numbers of days of the last 7 days you were active for
936 hs02$Q139 <- as.factor(hs02$Q139)
937 hs02$Q139 <- as.numeric(as.character(hs02$Q139))
938 hs02$Q139[hs02$Q139 == 8] <- 0 # 8 = no activity
939
940 # recode NA into 0 minute/hour
941 hs02$Q134_hrs[is.na(hs02$Q134_hrs)] <- 0

```

```

942 hs02$Q134_mins[is.na(hs02$Q134_mins)] <- 0
943 hs02$Q136_hrs[is.na(hs02$Q136_hrs)] <- 0
944 hs02$Q136_mins[is.na(hs02$Q136_mins)] <- 0
945 hs02$Q138_hrs[is.na(hs02$Q138_hrs)] <- 0
946 hs02$Q138_mins[is.na(hs02$Q138_mins)] <- 0
947
948 # calculate total physical activity duration per week in minutes (rigorous activity received double
    weight)
949 hs02$minutes <- (60*(hs02$Q134_hrs + hs02$Q136_hrs + 2*(hs02$Q138_hrs)) +
950                 (hs02$Q134_mins + hs02$Q136_mins + (2*hs02$Q138_mins)))
951
952 # recode them into active or not active
953 hs02$active <- ifelse(hs02$Q139 >= 5 & hs02$minutes >=30, "Active", "Not Active")
954 hs02$active <- as.factor(hs02$active)
955
956 hs02$active <- factor(hs02$active, levels=c("Not Active", "Active"))
957
958 prop_frame(hs02, "active")
959
960
961 ## Sedentary Lifestyle
962 # recode physical activity in minutes into sedentary or not sedentary
963 hs02$sedentary <- ifelse(hs02$minutes >=30, "Not Sedentary", "Sedentary" )
964
965 prop_frame(hs02, "sedentary")
966
967
968 ## Difficulty Climbing Several Steps of Stairs
969 hs02$Q234 <- as.factor(hs02$Q234)
970
971 hs02$Q234[hs02$Q234 == 4] <- NA # 4 = do not know
972 hs02$Q234[hs02$Q234 == 5] <- NA # 5 = refused
973 hs02$stair <- factor(hs02$Q234, labels=c("A Lot Difficult", "A Little Difficult", "No Difficulty"))
974
975 prop_frame(hs02, "stair")
976
977
978 ## Fruit and Vegetable Guideline
979 ## Q156= fruit , Q157= vegetable
980 hs02$Q156 <- as.factor(hs02$Q156)

```

```

981 hs02$Q157 <- as.factor(hs02$Q157)
982
983 summary(hs02$Q156)
984 summary(hs02$Q157)
985
986 hs02$Q156[hs02$Q156 == 11] <- NA # 12 = refused
987 hs02$Q156[hs02$Q156 == 12] <- NA # 12 = refused
988
989 hs02$Q157[hs02$Q157 == 11] <- NA
990 hs02$Q157[hs02$Q157 == 12] <- NA
991
992 # recode into meeting the serving suggestions of 2+/day for fruit and 3+/day for vegetable
993 hs02$fruit <- recode(hs02$Q156, "9"="No", "10"="No", "1"="No",
994                      .default="Yes")
995 hs02$fruit <- as.factor(hs02$fruit)
996
997 hs02$veges <- recode(hs02$Q157, "3"="Yes", "4"="Yes", .default="No")
998 hs02$veges <- as.factor(hs02$veges)
999
1000 prop_frame(hs02, "fruit")
1001 prop_frame(hs02, "veges")
1002
1003
1004 ## Rename cluster and strata so that they will merge with the rest of the NZHS
1005 names(hs02)[names(hs02) == 'PSU_no'] <- 'cluster'
1006 names(hs02)[names(hs02) == 'Stratum_no'] <- 'strata'
1007 names(hs02)[names(hs02) == 'Finalwgt'] <- 'finalwgt'
1008
1009
1010
1011 ##### NZHS 2006/07 – adult #####
1012
1013 ## Transform BMI into numeric, removing character strings
1014 summary(HS06A$BMI) # 407 Unknown (U) and 401 Refused (R)
1015 colnames(HS06A)[which(names(HS06A) == "BMI")] <- "bmi"
1016 HS06A$bmiscale <- as.numeric(as.character(HS06A$bmi))
1017 summary(HS06A$bmiscale)
1018
1019
1020 ## Recode 95+ in age variable into 95 (the max value is censored at this number)

```

```

1021 HS06A$age <- as.character(HS06A$age)
1022 HS06A$age[HS06A$age == "95+"] <- 95
1023 HS06A$age <- as.numeric(HS06A$age)
1024
1025 HS06A$D_01 <- as.factor(HS06A$D_01) # 1=male, 8=female
1026
1027 ## Recode BMI category using IOTF cut-offs
1028 # 0 = underweight; 1= average; 2= overweight; 3=obese
1029 HS06A <- HS06A %>%
1030   mutate(bmic = ifelse(D_01 == 1 & age == 15 & bmiscale <16.98, 0,
1031     ifelse(D_01 == 1 & age == 16 & bmiscale <17.53, 0,
1032       ifelse(D_01 == 1 & age == 17 & bmiscale <18.04, 0,
1033         ifelse(D_01 == 8 & age == 15 & bmiscale <17.43, 0,
1034           ifelse(D_01 == 8 & age == 16 & bmiscale <17.9, 0,
1035             ifelse(D_01 == 8 & age == 17 & bmiscale <18.24, 0,
1036               ifelse(age >= 18 & bmiscale <18.5, 0,
1037                 ifelse(D_01 == 1 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
1038                   ifelse(D_01 == 1 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
1039                     ifelse(D_01 == 1 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
1040                       ifelse(D_01 == 8 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
1041                         ifelse(D_01 == 8 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
1042                           ifelse(D_01 == 8 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
1043                             ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
1044                               ifelse(D_01 == 1 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
1045                                 ifelse(D_01 == 1 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
1046                                   ifelse(D_01 == 1 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
1047                                     ifelse(D_01 == 8 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
1048                                       ifelse(D_01 == 8 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
1049                                         ifelse(D_01 == 8 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
1050                                           ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
1051                                             ifelse(D_01 == 1 & age == 15 & bmiscale >=28.32, 3,
1052                                               ifelse(D_01 == 1 & age == 16 & bmiscale >=28.89, 3,
1053                                                 ifelse(D_01 == 1 & age == 17 & bmiscale >=29.43, 3,
1054                                                   ifelse(D_01 == 8 & age == 15 & bmiscale >=29.01, 3,
1055                                                     ifelse(D_01 == 8 & age == 16 & bmiscale >=29.40, 3,
1056                                                       ifelse(D_01 == 8 & age == 17 & bmiscale >=29.70, 3,
1057                                                         ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))
1058
1059 HS06A$bmic <- as.factor(HS06A$bmic)
1060 summary(HS06A$bmic) # 139 underweight, 808 NA

```



```

1061
1062
1063 # Subset the data, remove underweight
1064 a06 <- subset(HS06A, bmic != 0 | is.na(bmic))
1065
1066 # label
1067 a06$bmic <- factor(a06$bmic, labels=c("Average", "Overweight", "Obese"))
1068 summary(a06$bmic)
1069
1070
1071 ## Sex (D_01) (1 = male, 8 = female)
1072 sum(is.na(a06$D_01))
1073 a06$gender <- factor(a06$D_01, labels = c("Male", "Female"))
1074
1075 prop_frame(a06, "gender")
1076
1077
1078 ## Age
1079 summary(a06$age)
1080
1081
1082 ## Migration Status
1083 summary(a06$a5_05)
1084
1085 a06$native <- recode(a06$a5_05, "1997"="Migrant", " 1998" = "Migrant", "1999"="Migrant",
1086                             " 2000"="Migrant", "2001"="Migrant", "2002"="Migrant", " 2003"="Migrant",
1087                             "2004"="Migrant", "2005"="Migrant", "2006"="Migrant", "2007"="Migrant",
1088                             "K"= "No", "R"= "No", .default="Native") # K and R response are treated as NA
1089
1090 a06$native[is.na(a06$native)] <- "Native"
1091 a06$native[a06$native == "No"] <- NA
1092 a06$native <- as.factor(a06$native)
1093
1094 prop_frame(a06, "native")
1095
1096
1097 ## BMI
1098 summary(a06$bmiscale)
1099
1100 prop_frame(a06, "bmic")

```

```

1101
1102
1103 ## Ethnicity
1104 a06$A5_02_group_1 <- as.factor(a06$A5_02_group_1)
1105 a06$A5_02_group_2 <- as.factor(a06$A5_02_group_2)
1106 a06$A5_02_group_3 <- as.factor(a06$A5_02_group_3)
1107 a06$A5_02_group_4 <- as.factor(a06$A5_02_group_4)
1108 a06$A5_02_group_5 <- as.factor(a06$A5_02_group_5)
1109
1110
1111 # Recode anyone ~ European
1112 a06 <- a06 %>%
1113   mutate(euro = ifelse(A5_02_group_1 %in% c(1,9), 1,
1114     ifelse(A5_02_group_2 %in% c(1,9), 1,
1115       ifelse(A5_02_group_3 %in% c(1,9), 1, 0))))
1116
1117 # Recode anyone ~ Maori
1118 a06 <- a06 %>%
1119   mutate(maori = ifelse(A5_02_group_1 %in% 2, 1,
1120     ifelse(A5_02_group_2 %in% 2, 1,
1121       ifelse(A5_02_group_3 %in% 2, 1,
1122         ifelse(A5_02_group_4 %in% 2, 1, 0))))))
1123
1124 # Recode anyone ~ Pacific
1125 a06 <- a06 %>%
1126   mutate(pacific = ifelse(A5_02_group_1 %in% c(3,4,5,6,10), 1,
1127     ifelse(A5_02_group_2 %in% c(3,4,5,6,10), 1,
1128       ifelse(A5_02_group_3 %in% c(3,4,5,6,10), 1,
1129         ifelse(A5_02_group_4 %in% c(3,4,5,6,10), 1,
1130           ifelse(A5_02_group_5 %in% c(3,4,5,6,10), 1, 0)))))))
1131
1132 # Recode anyone ~ Asian (Indian and Chinese only)
1133 a06 <- a06 %>%
1134   mutate(asian = ifelse(A5_02_group_1 %in% c(7,8), 1,
1135     ifelse(A5_02_group_2 %in% c(7,8), 1,
1136       ifelse(A5_02_group_3 %in% c(7,8), 1,
1137         ifelse(A5_02_group_4 %in% c(7,8), 1,
1138           ifelse(A5_02_group_5 %in% c(7,8), 1, 0)))))))
1139
1140 # Recode anyone ~ Other (other asian is here)

```

```

1141 a06 <- a06 %>%
1142   mutate(other = ifelse(A5_02_group_1 %in% c(11,12), 1,
1143     ifelse(A5_02_group_2 %in% c(11,12), 1,
1144       ifelse(A5_02_group_3 %in% c(11,12), 1,
1145         ifelse(A5_02_group_4 %in% c(11,12), 1,
1146           ifelse(A5_02_group_5 %in% c(11,12), 1, 0))))))
1147
1148
1149 # Recode them into: Maori only, Pacific only, Asian only*, European only, Other, 2+ ethnicities (M)
1150 a06 = a06 %>%
1151   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
1152     Only",
1153       ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
1154         Pacific Only",
1155         ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
1156           Only",
1157           ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
1158             European Only",
1159             ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
1160               ,
1161               ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
1162                 Ethnicities (M)",
1163                 ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
1164                   Ethnicities (M)",
1165                   ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
1166                     ,
1167                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
1168                       ,
1169                       ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
1170                         ,
1171                         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
1172                           Ethnicities (M)",
1173                           ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
1174                             Ethnicities (M)",
1175                           ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
1176                             ,
1177                             ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
1178                               ,
1179                               ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
1180                                 ,
1181                                 )
1182                             )
1183                           )
1184                         )
1185                       )
1186                     )
1187                   )
1188                 )
1189               )
1190             )
1191           )
1192         )
1193       )
1194     )

```

```

1166         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
           Ethnicities (M)",
1167         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
           Ethnicities (M)",
1168         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
           Ethnicities (M)",
1169         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
           ,
1170         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
           ,
1171         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
           ,
1172         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
           Ethnicities (M)",
1173         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
           Ethnicities (M)",
1174         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
           Ethnicities (M)",
1175         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
           ,
1176         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
           Ethnicities (M)",
1177         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
           Ethnicities (M)",
1178         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
           Ethnicities (M)",
1179         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
           ,
1180         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
           Ethnicities (M)",
1181         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
           Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))))))
1182
1183 a06$eth_count <- as.factor(a06$eth_count)
1184
1185 prop_frame(a06, "eth_count")
1186
1187
1188 ## Deprivation Quintile
1189 a06$nzdep06_quintile <- as.factor(a06$nzdep06_quintile)

```

```

1190
1191 names(a06)[names(a06) == "nzdep06_quintile"] <- 'dep'
1192
1193 prop_frame(a06, "dep")
1194
1195
1196 ## Fruit and Vegetable
1197 # 28 = fruit , 29 = vegetable
1198 a06$A3_28 <- as.factor(a06$A3_28)
1199 a06$A3_29 <- as.factor(a06$A3_29)
1200 summary(a06$A3_28)
1201 summary(a06$A3_29) # 1 answered "don't know"
1202
1203 a06$A3_29[a06$A3_29 == "K"] <- NA
1204
1205 # Adherence to fruit and vegetable guideline
1206 a06$fruit <- as.numeric(as.character(a06$A3_28))
1207 a06$fruit <- cut(a06$fruit , breaks=c(0,3,6))
1208 summary(a06$fruit)
1209
1210 a06$fruit <- factor(a06$fruit , labels=c("No" ,"Yes"))
1211
1212 a06$veges <- as.numeric(as.character(a06$A3_28))
1213 a06$veges <- cut(a06$veges , breaks=c(0,4,6))
1214 summary(a06$veges)
1215
1216 a06$veges <- factor(a06$veges , labels=c("No" ,"Yes"))
1217
1218 prop_frame(a06, "fruit")
1219 prop_frame(a06, "veges")
1220
1221
1222 ## Smoking Status
1223 a06 <- a06 %>%
1224   mutate(smoke = ifelse(ex_smoker == 0 & current_smoker == 0, "Non Smoker",
1225     ifelse(ex_smoker == 1 & current_smoker == 0, "Ex Smoker", "Current Smoker")))
1226
1227 a06$smoke <- as.factor(a06$smoke)
1228 a06$smoke <- factor(a06$smoke, levels = c("Non Smoker", "Ex Smoker", "Current Smoker"))
1229

```

```

1230 prop_frame(a06, "smoke")
1231
1232
1233 ## Alcohol Problem
1234 a06$AUDIT_cat[a06$AUDIT_cat == "K"] <- NA
1235 a06$AUDIT_cat[a06$AUDIT_cat == "R"] <- NA
1236
1237 a06$haz_drinker_all <- recode(a06$AUDIT_cat, "0"="No Alcohol Problem", "1"="Alcohol Problem", .
    default= NULL)
1238
1239 prop_frame(a06, "haz_drinker_all")
1240
1241
1242 ## Urban/Rural Area
1243 a06$UA <- as.factor(a06$UR_06)
1244 a06$UA <- recode(a06$UA, "4"="Rural", .default = "Urban")
1245 a06$UA <- as.factor(a06$UA)
1246
1247 prop_frame(a06, "UA")
1248
1249
1250 ## Difficulty Climbing Several Steps of Stair
1251 a06$stair <- as.factor(a06$A4_22)
1252 a06$stair[a06$stair == "K"] <- NA
1253
1254 a06$stair <- factor(a06$stair, labels=c("A Lot Difficult","A Little Difficult","No Difficulty"))
1255
1256 prop_frame(a06, "stair")
1257
1258
1259 ## Physical Activity
1260 a06$A3_12 <- as.numeric(as.character(a06$A3_12))
1261 a06$A3_14 <- as.numeric(as.character(a06$A3_14))
1262 a06$A3_16 <- as.numeric(as.character(a06$A3_16))
1263 a06$A3_13a <- as.numeric(as.character(a06$A3_13a))
1264 a06$A3_15a <- as.numeric(as.character(a06$A3_15a))
1265 a06$A3_17a <- as.numeric(as.character(a06$A3_17a))
1266 a06$A3_13b <- as.numeric(as.character(a06$A3_13b))
1267 a06$A3_15b <- as.numeric(as.character(a06$A3_15b))
1268 a06$A3_17b <- as.numeric(as.character(a06$A3_17b))

```

```

1269 a06 <- mutate(a06, minutes= (60*(A3_13a+A3_15a+2*(A3_17a)))+A3_13b+A3_15b+(2*A3_17b))
1270 summary(a06$minutes) # physical activity in minutes in the past week
1271
1272 # active at least 30' per day for 5 days in the last 1 week (set question)
1273 a06$active <- as.numeric(as.character(a06$A3_18))
1274 a06$active <- cut(a06$active, breaks=c(-Inf,4,Inf), labels=c("Not Active","Active"))
1275
1276 prop_frame(a06, "active")
1277
1278 ## Sedentary Lifestyle
1279 a06 <- mutate(a06, sedentary = ifelse(minutes <30, "Sedentary", "Not Sedentary"))
1280
1281 a06$sedentary <- as.factor(a06$sedentary)
1282
1283 prop_frame(a06, "sedentary")
1284
1285
1286 ## Household Income
1287 a06$hhinc <- ordered(a06$A5_24, levels=c("1","2","3","4","5","6","7","8","9","10","11","12","13","14",
1288     "15"))
1289 sum(is.na(a06$hhinc)) # 1450 missing
1290 sum(is.na(a06$hhinc))/length(a06$hhinc) # 12.5639% missing
1291
1292 a06$hhinc <- as.numeric(as.character(a06$hhinc))
1293 a06$hhinc <- cut(a06$hhinc, breaks=c(0,3,7,10,15))
1294 a06$hhinc <- factor(a06$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000", ">$70,000"))
1295
1296 prop_frame(a06, "hhinc")
1297
1298 ## Education
1299 a06$secondary <- recode(a06$A5_13, "1" = "1", .default = "2") # 1= no degree, 2= secondary school
1300     grads
1301
1302 a06$tertiary <- recode(a06$A5_14_group, "1" = "0", .default = "3") # 0 = no tertiary, 3= tertiary
1303
1304 a06 <- mutate(a06, edu = ifelse(secondary == "1" & tertiary == "0", 1,
1305     ifelse(secondary == "2" & tertiary == "0", 2,
1306     ifelse(tertiary == "3", 3, 1))))

```

```

1307 a06$edu <- as.factor(a06$edu)
1308 a06$edu <- factor(a06$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
1309
1310 prop_frame(a06, "edu")
1311
1312
1313 ## Recode unique clusters for each stratum
1314 names(a06)[names(a06) == "stratum"] <- "strata"
1315
1316 summary(a06$cluster) # min 1, max 137; start recoding with 138
1317
1318 # create a code book for a new cluster coding
1319 l <- c(1:137) # Var1 = cluster
1320 m <- c(1:21) # Var2 = strata
1321
1322 k <- data.frame(expand.grid(l, m))
1323 k$new.cluster <- c(1:nrow(k))
1324
1325 # create a backup of original clusters
1326 a06$clu <- a06$cluster
1327
1328 # recode them into new unique clusters
1329 for (i in 1:137){
1330   for (j in 1:21){
1331     a06$cluster[a06$cluster == i & a06$strata == j] <- k$new.cluster[k$Var1 == i & k$Var2 == j]
1332   }
1333 }
1334
1335 # save object "k" to use it in the child data set
1336
1337
1338
1339 ##### NZHS 2006/07 – child #####
1340
1341 summary(HS06C$BMI) # 148 refused, 8 unknown
1342 colnames(HS06C)[which(names(HS06C) == "BMI")] <- "bmi"
1343
1344 # remove non-numeric data
1345 HS06C$bmiscale <- (as.numeric(as.character(HS06C$bmi)))
1346

```



```

1347 HS06C <- HS06C %>%
1348   mutate(under =
1349     ifelse(sex == "M" & age == 2 & bmiscale <15.24, 0,
1350     ifelse(sex == "M" & age == 3 & bmiscale <14.83, 0,
1351     ifelse(sex == "M" & age == 4 & bmiscale <14.51, 0,
1352     ifelse(sex == "M" & age == 5 & bmiscale <14.26, 0,
1353     ifelse(sex == "M" & age == 6 & bmiscale <14.06, 0,
1354     ifelse(sex == "M" & age == 7 & bmiscale <14.00, 0,
1355     ifelse(sex == "M" & age == 8 & bmiscale <14.13, 0,
1356     ifelse(sex == "M" & age == 9 & bmiscale <14.36, 0,
1357     ifelse(sex == "M" & age == 10 & bmiscale <14.63, 0,
1358     ifelse(sex == "M" & age == 11 & bmiscale <14.96, 0,
1359     ifelse(sex == "M" & age == 12 & bmiscale <15.36, 0,
1360     ifelse(sex == "M" & age == 13 & bmiscale <15.84, 0,
1361     ifelse(sex == "M" & age == 14 & bmiscale <16.39, 0,
1362     ifelse(sex == "M" & age == 15 & bmiscale <16.98, 0,
1363     ifelse(sex == "M" & age == 16 & bmiscale <17.53, 0,
1364     ifelse(sex == "M" & age == 17 & bmiscale <18.04, 0,
1365     ifelse(sex == "F" & age == 2 & bmiscale <14.96, 0,
1366     ifelse(sex == "F" & age == 3 & bmiscale <14.60, 0,
1367     ifelse(sex == "F" & age == 4 & bmiscale <14.30, 0,
1368     ifelse(sex == "F" & age == 5 & bmiscale <14.04, 0,
1369     ifelse(sex == "F" & age == 6 & bmiscale <13.85, 0,
1370     ifelse(sex == "F" & age == 7 & bmiscale <13.83, 0,
1371     ifelse(sex == "F" & age == 8 & bmiscale <14.00, 0,
1372     ifelse(sex == "F" & age == 9 & bmiscale <14.26, 0,
1373     ifelse(sex == "F" & age == 10 & bmiscale <14.58, 0,
1374     ifelse(sex == "F" & age == 11 & bmiscale <15.03, 0,
1375     ifelse(sex == "F" & age == 12 & bmiscale <15.59, 0,
1376     ifelse(sex == "F" & age == 13 & bmiscale <16.23, 0,
1377     ifelse(sex == "F" & age == 14 & bmiscale <16.86, 0,
1378     ifelse(sex == "F" & age == 15 & bmiscale <17.43, 0,
1379     ifelse(sex == "F" & age == 16 & bmiscale <17.90, 0,
1380     ifelse(sex == "F" & age == 17 & bmiscale <18.24, 0,
1381     ifelse(age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))
1382
1383 HS06C <- HS06C %>%
1384   mutate(average =
1385     ifelse(sex == "M" & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
1386     ifelse(sex == "M" & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,

```

```

1387   ifelse (sex == "M" & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
1388   ifelse (sex == "M" & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
1389   ifelse (sex == "M" & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
1390   ifelse (sex == "M" & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
1391   ifelse (sex == "M" & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
1392   ifelse (sex == "M" & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
1393   ifelse (sex == "M" & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
1394   ifelse (sex == "M" & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
1395   ifelse (sex == "M" & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
1396   ifelse (sex == "M" & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
1397   ifelse (sex == "M" & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
1398   ifelse (sex == "M" & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
1399   ifelse (sex == "M" & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
1400   ifelse (sex == "M" & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
1401   ifelse (sex == "F" & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
1402   ifelse (sex == "F" & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
1403   ifelse (sex == "F" & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
1404   ifelse (sex == "F" & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
1405   ifelse (sex == "F" & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
1406   ifelse (sex == "F" & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
1407   ifelse (sex == "F" & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
1408   ifelse (sex == "F" & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
1409   ifelse (sex == "F" & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
1410   ifelse (sex == "F" & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
1411   ifelse (sex == "F" & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
1412   ifelse (sex == "F" & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
1413   ifelse (sex == "F" & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
1414   ifelse (sex == "F" & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
1415   ifelse (sex == "F" & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
1416   ifelse (sex == "F" & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
1417   ifelse (age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA))))))))))))))))))))))))))))))
1418
1419 HS06C <- HS06C %>%
1420   mutate (over=
1421     ifelse (sex == "M" & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
1422     ifelse (sex == "M" & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
1423     ifelse (sex == "M" & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
1424     ifelse (sex == "M" & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
1425     ifelse (sex == "M" & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
1426     ifelse (sex == "M" & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,

```

```

1427 ifelse (sex == "M" & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
1428 ifelse (sex == "M" & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
1429 ifelse (sex == "M" & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
1430 ifelse (sex == "M" & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
1431 ifelse (sex == "M" & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
1432 ifelse (sex == "M" & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
1433 ifelse (sex == "M" & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
1434 ifelse (sex == "M" & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
1435 ifelse (sex == "M" & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
1436 ifelse (sex == "M" & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
1437 ifelse (sex == "F" & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
1438 ifelse (sex == "F" & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
1439 ifelse (sex == "F" & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
1440 ifelse (sex == "F" & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
1441 ifelse (sex == "F" & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
1442 ifelse (sex == "F" & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
1443 ifelse (sex == "F" & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
1444 ifelse (sex == "F" & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
1445 ifelse (sex == "F" & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
1446 ifelse (sex == "F" & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
1447 ifelse (sex == "F" & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
1448 ifelse (sex == "F" & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
1449 ifelse (sex == "F" & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
1450 ifelse (sex == "F" & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
1451 ifelse (sex == "F" & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
1452 ifelse (sex == "F" & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
1453 ifelse (age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))
1454
1455 HS06C <- HS06C %>%
1456 mutate (obes=
1457 ifelse (sex == "M" & age == 2 & bmiscale >=19.99, 3,
1458 ifelse (sex == "M" & age == 3 & bmiscale >=19.50, 3,
1459 ifelse (sex == "M" & age == 4 & bmiscale >=19.23, 3,
1460 ifelse (sex == "M" & age == 5 & bmiscale >=19.27, 3,
1461 ifelse (sex == "M" & age == 6 & bmiscale >=19.76, 3,
1462 ifelse (sex == "M" & age == 7 & bmiscale >=20.59, 3,
1463 ifelse (sex == "M" & age == 8 & bmiscale >=21.56, 3,
1464 ifelse (sex == "M" & age == 9 & bmiscale >=22.71, 3,
1465 ifelse (sex == "M" & age == 10 & bmiscale >=23.96, 3,
1466 ifelse (sex == "M" & age == 11 & bmiscale >=25.07, 3,

```

```

1467 ifelse (sex == "M" & age == 12 & bmiscale >=26.02, 3,
1468 ifelse (sex == "M" & age == 13 & bmiscale >=26.87, 3,
1469 ifelse (sex == "M" & age == 14 & bmiscale >=27.64, 3,
1470 ifelse (sex == "M" & age == 15 & bmiscale >=28.32, 3,
1471 ifelse (sex == "M" & age == 16 & bmiscale >=28.89, 3,
1472 ifelse (sex == "M" & age == 17 & bmiscale >=29.43, 3,
1473 ifelse (sex == "F" & age == 2 & bmiscale >=19.81, 3,
1474 ifelse (sex == "F" & age == 3 & bmiscale >=19.38, 3,
1475 ifelse (sex == "F" & age == 4 & bmiscale >=19.16, 3,
1476 ifelse (sex == "F" & age == 5 & bmiscale >=19.20, 3,
1477 ifelse (sex == "F" & age == 6 & bmiscale >=19.61, 3,
1478 ifelse (sex == "F" & age == 7 & bmiscale >=20.39, 3,
1479 ifelse (sex == "F" & age == 8 & bmiscale >=21.44, 3,
1480 ifelse (sex == "F" & age == 9 & bmiscale >=22.66, 3,
1481 ifelse (sex == "F" & age == 10 & bmiscale >=23.97, 3,
1482 ifelse (sex == "F" & age == 11 & bmiscale >=25.25, 3,
1483 ifelse (sex == "F" & age == 12 & bmiscale >=26.47, 3,
1484 ifelse (sex == "F" & age == 13 & bmiscale >=27.57, 3,
1485 ifelse (sex == "F" & age == 14 & bmiscale >=28.42, 3,
1486 ifelse (sex == "F" & age == 15 & bmiscale >=29.01, 3,
1487 ifelse (sex == "F" & age == 16 & bmiscale >=29.40, 3,
1488 ifelse (sex == "F" & age == 17 & bmiscale >=29.70, 3,
1489 ifelse (age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))
1490
1491
1492 HS06C <- HS06C %>%
1493   mutate(bmic = ifelse (under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
1494     ifelse (average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
1495     ifelse (over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
1496     ifelse (obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA))))
1497
1498 HS06C$bmic <- as.factor(HS06C$bmic)
1499 summary(HS06C$bmic) # 134 underweight, 903 NA
1500
1501
1502 # exclude below 2 years old and underweight
1503 c06 <- subset(HS06C, age>=2 & bmic != 0 | is.na(bmic))
1504
1505
1506 ## BMI

```

```

1507 c06$bmic <- factor(c06$bmic, labels=c("Average", "Overweight", "Obese"))
1508
1509 prop_frame(c06, "bmic")
1510
1511
1512 ## Sex
1513 c06$gender <- factor(c06$sex, labels = c("Female", "Male"))
1514 c06$gender <- factor(c06$gender, levels = c("Male", "Female"))
1515
1516 prop_frame(c06, "gender")
1517
1518
1519 ## Age
1520 summary(c06$age)
1521
1522
1523 ## Ethnicity
1524 # recode anyone ~ european
1525 c06 <- c06 %>%
1526   mutate(euro = ifelse(C4_02_group_1 %in% c(1,9), 1,
1527     ifelse(C4_02_group_2 %in% c(1,9), 1,
1528       ifelse(C4_02_group_3 %in% c(1,9), 1, 0))))
1529
1530 # recode anyone ~ maori
1531 c06 <- c06 %>%
1532   mutate(maori = ifelse(C4_02_group_1 %in% 2, 1,
1533     ifelse(C4_02_group_2 %in% 2, 1,
1534       ifelse(C4_02_group_3 %in% 2, 1,
1535         ifelse(C4_02_group_4 %in% 2, 1, 0))))))
1536
1537 # recode anyone ~ Pacific
1538 c06 <- c06 %>%
1539   mutate(pacific = ifelse(C4_02_group_1 %in% c(3,4,5,6,10), 1,
1540     ifelse(C4_02_group_2 %in% c(3,4,5,6,10), 1,
1541       ifelse(C4_02_group_3 %in% c(3,4,5,6,10), 1,
1542         ifelse(C4_02_group_4 %in% c(3,4,5,6,10), 1,
1543           ifelse(C4_02_group_5 %in% c(3,4,5,6,10), 1,
1544             ifelse(C4_02_group_6 %in% c(3,4,5,6,10), 1,
1545               ifelse(C4_02_group_7 %in% c(3,4,5,6,10), 1, 0))))))))))
1546

```

```

1547 # recode anyone ~ Asian (Indian and Chinese only)
1548 c06 <- c06 %>%
1549   mutate(asian = ifelse(C4_02_group_1 %in% c(7,8), 1,
1550     ifelse(C4_02_group_2 %in% c(7,8), 1,
1551       ifelse(C4_02_group_3 %in% c(7,8), 1,
1552         ifelse(C4_02_group_4 %in% c(7,8), 1,
1553           ifelse(C4_02_group_5 %in% c(7,8), 1,
1554             ifelse(C4_02_group_6 %in% c(7,8), 1,
1555               ifelse(C4_02_group_7 %in% c(7,8), 1, 0))))))))))
1556
1557 # recode anyone ~ Other (other asian is here)
1558 c06 <- c06 %>%
1559   mutate(other = ifelse(C4_02_group_1 %in% c(11,12), 1,
1560     ifelse(C4_02_group_2 %in% c(11,12), 1,
1561       ifelse(C4_02_group_3 %in% c(11,12), 1,
1562         ifelse(C4_02_group_4 %in% c(11,12), 1,
1563           ifelse(C4_02_group_5 %in% c(11,12), 1,
1564             ifelse(C4_02_group_6 %in% c(11,12), 1,
1565               ifelse(C4_02_group_7 %in% c(11,12), 1, 0))))))))))
1566
1567
1568 # Recode them into: Maori only, Pacific only, Asian only, European only, Other, 2+ ethnicities (M)
1569 c06 = c06 %>%
1570   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
1571     Only",
1572     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
1573       Pacific Only",
1574     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
1575       Only",
1576     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
1577       European Only",
1578     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
1579       ,
1580     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
1581       Ethnicities (M)",
1582     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
1583       Ethnicities (M)",
1584     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
1585       ,

```

```

1578         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
1579         ,
1580         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
1581         ,
1582         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
1583             Ethnicities (M)",
1584         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
1585             Ethnicities (M)",
1586         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
1587         ,
1588         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
1589         ,
1590         ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
1591         ,
1592         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
1593             Ethnicities (M)",
1594         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
1595             Ethnicities (M)",
1596         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
1597             Ethnicities (M)",
1598         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
1599         ,
1600         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
1601         ,
1602         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
1603         ,
1604         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
1605             Ethnicities (M)",
1606         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
1607             Ethnicities (M)",
1608         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
1609             Ethnicities (M)",
1610         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
1611         ,
1612         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
1613             Ethnicities (M)",
1614         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
1615             Ethnicities (M)",
1616         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
1617             Ethnicities (M)",

```

```

1598         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
1599         ,
1600         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
            Ethnicities (M)",
1601         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
            Ethnicities (M)", NA))))))))))))))))))))))))))))))))))
1602 c06$eth_count <- as.factor(c06$eth_count)
1603
1604 prop_frame(c06, "eth_count")
1605
1606
1607 ## Urban/Rural Area
1608 c06$UA <- recode(c06$ur_desc, "Rural"="Rural", .default="Urban")
1609
1610 prop_frame(c06, "UA")
1611
1612
1613 ## Household Income
1614 summary(c06$C4_22)
1615 c06$hhinc <- ordered(c06$C4_22, levels=c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12",
1616         "13", "14", "15"))
1617 sum(is.na(c06$hhinc))
1618 sum(is.na(c06$hhinc))/length(c06$hhinc) # 369 NA, 9.5% missing
1619
1620 c06$hhinc <- as.numeric(as.character(c06$hhinc))
1621 c06$hhinc <- cut(c06$hhinc, breaks=c(0,3,7,10,15))
1622 c06$hhinc <- factor(c06$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000", ">$70,000"))
1623
1624 prop_frame(c06, "hhinc")
1625
1626
1627 ## Education
1628 c06$edu1 <- as.numeric(as.character(c06$C4_19))
1629 c06$secondary <- recode(c06$edu1, "1" = "1", .default = "2") # 1= no degree, 2= secondary school
            grads
1630
1631 summary(c06$C4_20_group)
1632 c06$edu2 <- as.numeric(as.character(c06$C4_20_group))
1633 c06$tertiary <- recode(c06$edu2, "1" = "0", .default = "3") #0 = no tertiary, 3= tertiary

```



```

1634
1635 # prioritise highest educational qualification in the household
1636 c06 <- mutate(c06, edu = ifelse(secondary == "1" & tertiary == "0", 1,
1637                                ifelse(secondary == "2" & tertiary == "0", 2,
1638                                ifelse(tertiary == "3", 3, 1)))
1639
1640 c06$edu <- as.factor(c06$edu)
1641 c06$edu <- factor(c06$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
1642
1643 prop_frame(c06, "edu")
1644
1645
1646 ## Deprivation Quintile
1647 names(c06)[names(c06) == "nzdep06_quintile"] <- "dep"
1648
1649 prop_frame(c06, "dep")
1650
1651
1652 ## Soft Drink and Fast Food
1653 # categorise the soft drink consumption to (0 = 0/week, 1 = 1/week, 2 = 2-3/week, 3 = 4+/week)
1654 c06$C3_13 <- as.factor(c06$C3_13)
1655 c06$softd <- c06$C3_13
1656 c06$softd <- recode(c06$softd, "0"="0/week", "1"="1/week", "2" = "2-3/week", "3" = "2-3/week",
1657                        .default="4+/week")
1658 c06$softd <- ordered(c06$softd, levels = c("0/week", "1/week", "2-3/week", "4+/week"))
1659
1660 prop_frame(c06, "softd")
1661
1662
1663 # categorise fast food consumption to (0 = 0/week, 1 = 1/week, 2 = 2-3/week, 3 = 4+/week)
1664 c06$fastf <- c06$C3_15
1665 c06$fastf <- recode(c06$fastf, "0"="0/week", "1"="1/week", "2"="2-3/week",
1666                        "3"="2-3/week", .default="4+/week")
1667 c06$fastf <- ordered(c06$fastf, levels =c("0/week", "1/week", "2-3/week", "4+/week"))
1668
1669 prop_frame(c06, "fastf")
1670
1671
1672 ## Recode Unique Clusters
1673 names(c06)[names(c06) == "stratum"] <- "strata"

```

```

1674
1675 # create backup
1676 c06$clu <- c06$cluster
1677
1678 # recode to unique clusters using the same codebook (k) as the one in adult data set
1679 for (i in 1:131){
1680   for (j in 1:21){
1681     c06$cluster[c06$cluster == i & c06$strata == j] <- k$new.cluster[k$Var1 == i & k$Var2 == j]
1682   }
1683 }
1684
1685 summary(c06$cluster)
1686
1687 # remove codebook "k" from the memory
1688 rm(k)
1689
1690
1691 ##### NZHS 2011/12 – adult #####
1692
1693 # R= refused, U= unable, P= pregnant, O=outliers
1694 summary(HS11A$bmi) # 941 R, 637 U, 302 P, 20 O
1695
1696 HS11A$bmiscale <- as.numeric(as.character(HS11A$bmi))
1697 summary(HS11A$bmiscale) # 1900 converted into NA
1698
1699 HS11A$gender <- as.factor(HS11A$gender) # 0 = male, 1 = female
1700
1701 HS11A <- HS11A %>%
1702   mutate(bmic = ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
1703     ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
1704       ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
1705         ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
1706           ifelse(gender == 1 & age == 16 & bmiscale <17.9, 0,
1707             ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
1708               ifelse(age >= 18 & bmiscale <18.5, 0,
1709                 ifelse(gender == 0 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
1710                   ifelse(gender == 0 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
1711                     ifelse(gender == 0 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
1712                       ifelse(gender == 1 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
1713                         ifelse(gender == 1 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,

```

```

1714         ifelse(gender == 1 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
1715         ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
1716         ifelse(gender == 0 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
1717         ifelse(gender == 0 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
1718         ifelse(gender == 0 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
1719         ifelse(gender == 1 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
1720         ifelse(gender == 1 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
1721         ifelse(gender == 1 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
1722         ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
1723         ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
1724         ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
1725         ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
1726         ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
1727         ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
1728         ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
1729         ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))
1730
1731 HS11A$bmic <- as.factor(HS11A$bmic)
1732 summary(HS11A$bmic)
1733
1734 # exclude underweight
1735 a11 <- subset(HS11A, !bmic == 0 | is.na(bmic))
1736
1737
1738 ## Migration Status
1739 # recode migrants <10–11 years into "Migrant", .K and .R into NA, and the rest into "Native"
1740 a11$native <- recode(a11$A5_06, "2002"="Migrant", "2003"="Migrant", "2004"="Migrant", "2005"="
1741         Migrant",
1742         "2006"="Migrant", "2007"="Migrant", "2008"="Migrant", "2009"="
1743         Migrant",
1744         "2010"="Migrant", "2011"="Migrant", ".K"= "S", ".R"= "S",
1745         .default="Native")
1746
1747 summary(a11$native)
1748 a11$native[a11$native == "S"] <- NA #recode S into NA
1749
1750 a11$native <- factor(a11$native, levels=c("Native", "Migrant"))
1751
1752 prop_frame(a11, "native")

```

```

1752
1753 ## BMI
1754 a11$bmic <- factor(a11$bmic, labels=c("Average", "Overweight", "Obese"))
1755
1756 prop_frame(a11, "bmic")
1757
1758 summary(a11$bmiscale)
1759
1760
1761 ## Sex
1762 a11$gender <- factor(a11$gender, labels=c("Male", "Female"))
1763
1764 prop_frame(a11, "gender")
1765
1766
1767 ## Ethnicity
1768 # recode anyone ~ european
1769 a11 = a11 %>%
1770   mutate(euro = ifelse(A5_03_01 == 1, 1, 0))
1771
1772 # recode anyone ~ maori
1773 a11 = a11 %>%
1774   mutate(maori = ifelse(A5_03_02 == 1, 1, 0))
1775
1776 # recode anyone ~ pacific
1777 a11 = a11 %>%
1778   mutate(pacific = ifelse(A5_03_03 == 1, 1,
1779     ifelse(A5_03_04 == 1, 1,
1780       ifelse(A5_03_05 == 1, 1,
1781         ifelse(A5_03_06 == 1, 1, 0))))))
1782
1783 # recode anyone ~ asian
1784 # starting from 2011, other asian will be coded as other ethnicity because there is no way to
1785   identify other asian
1786 a11 = a11 %>%
1787   mutate(asian = ifelse(A5_03_07 == 1, 1,
1788     ifelse(A5_03_08 == 1, 1, 0)))
1789
1790 # recode anyone ~ other
1791 a11 = a11 %>%

```

```

1791 mutate(other = ifelse(A5_03_77 == 1, 1, 0))
1792
1793
1794 a11 = a11 %>%
1795   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
1796     Only",
1797       ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
1798         Pacific Only",
1799       ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
1800         Only",
1801       ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
1802         European Only",
1803       ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
1804         ,
1805       ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
1806         Ethnicities (M)",
1807       ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
1808         Ethnicities (M)",
1809       ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
1810         ,
1811       ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
1812         ,
1813       ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
1814         ,
1815       ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
1816         Ethnicities (M)",
1817       ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
1818         Ethnicities (M)",
1819       ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
1820         ,
1821       ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
1822         ,
1823       ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
1824         ,
1825       ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
1826         Ethnicities (M)",
1827       ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
1828         Ethnicities (M)",
1829       ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
1830         Ethnicities (M)",

```

```

1813         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
1814         ,
1815         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
1816         ,
1817         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
1818         ,
1819         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
1820         Ethnicities (M)",
1821         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
1822         Ethnicities (M)",
1823         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
1824         Ethnicities (M)",
1825         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
1826         ,
1827         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
1828         Ethnicities (M)",
1829         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
1830         Ethnicities (M)",
1831         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
1832         Ethnicities (M)",
1833         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
1834         ,
1835         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
1836         Ethnicities (M)",
1837         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
1838         Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))))))
1839
1840 a11$eth_count = as.factor(a11$eth_count)
1841
1842 prop_frame(a11, "eth_count")
1843
1844
1845 ## Household Income
1846 a11$hhinc <- ordered(a11$A5_24, levels=c("4", "5", "6", "7", "8", "9", "10", "11", "12",
1847         "13", "14", "15", "16"))
1848
1849 sum(is.na(a11$hhinc))
1850 sum(is.na(a11$hhinc))/length(a11$hhinc) # 369 NA, 9.5% missing
1851
1852 a11$hhinc <- as.numeric(as.character(a11$hhinc))
1853 a11$hhinc <- cut(a11$hhinc, breaks=c(0,5,10,13,16))

```

```

1840 a11$hhinc <- factor(a11$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000", ">$70,000"))
1841
1842 prop_frame(a11, "hhinc")
1843
1844
1845 ## Education
1846 # 1= no qualification, 2= secondary school graduates, 3= tertiary school graduate
1847 summary(a11$A5_14)
1848 a11$secondary <- as.numeric(as.character(a11$A5_14))
1849 a11$secondary <- recode(a11$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
1850 a11$secondary <- as.factor(a11$secondary)
1851
1852 summary(a11$A5_15)
1853 a11$tertiary <- as.numeric(as.character(a11$A5_15))
1854 a11$tertiary <- recode(a11$tertiary, "0" = "0", "1"="2",
1855                        "2"="2", "3"="2", "4"="2", .default = "3") #0= no tertiary, 2= secondary, 3=
    tertiary
1856 a11$tertiary <- as.factor(a11$tertiary)
1857
1858 a11 <- mutate(a11, edu = ifelse(secondary == "1" & tertiary == "0", 1,
1859                                ifelse(secondary == "2" & tertiary == "0", 2,
1860                                ifelse(tertiary == "2", 2,
1861                                ifelse(tertiary == "3", 3, 1))))))
1862
1863 a11$edu <- as.factor(a11$edu)
1864 summary(a11$edu) #100 NA
1865 sum(is.na(a11$edu))/length(a11$edu) # 0.95 % missing
1866
1867 a11$edu <- factor(a11$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
1868
1869 prop_frame(a11, "edu")
1870
1871
1872 ## Deprivation Quintile
1873 names(a11)[names(a11) == "nzdep06_quintile"] <- "dep"
1874
1875 prop_frame(a11, "dep")
1876
1877

```

```

1878 ## Fruit and Vegetable
1879 summary(a11$fruit)
1880 summary(a11$veges)
1881 a11$fruit[a11$fruit == "K"] <- NA
1882 a11$fruit[a11$fruit == "R"] <- NA
1883 a11$veges[a11$veges == "K"] <- NA
1884 a11$veges[a11$veges == "R"] <- NA
1885 a11$fruit <- factor(a11$fruit, labels=c("No", "Yes"))
1886 a11$veges <- factor(a11$veges, labels=c("No", "Yes"))
1887
1888 prop_frame(a11, "fruit")
1889
1890 prop_frame(a11, "veges")
1891
1892
1893 ## Alcohol Problem
1894 a11$haz_drinker_all[a11$haz_drinker_all == "K"] <- NA
1895 a11$haz_drinker_all[a11$haz_drinker_all == "R"] <- NA
1896 a11$haz_drinker_all <- factor(a11$haz_drinker_all, labels = c("No Alcohol Problem", "Alcohol Problem"
    ))
1897
1898 prop_frame(a11, "haz_drinker_all")
1899
1900
1901 ## Smoking Status
1902 # checking for discrepancies
1903 a11 %>%
1904   filter(!ex_smoker == "K" & !ex_smoker == "R") %>%
1905   group_by(ex_smoker) %>%
1906   summarise(n=n()) %>%
1907   mutate(prop=n/sum(n))
1908
1909 a11 %>%
1910   filter(!never_tried_smoking == "K" & !never_tried_smoking == "R") %>%
1911   group_by(never_tried_smoking) %>%
1912   summarise(n=n()) %>%
1913   mutate(prop=n/sum(n))
1914
1915 a11 %>%
1916   filter(!current_smoker == "K" & !current_smoker == "R") %>%

```



```

1917 group_by(current_smoker) %>%
1918 summarise(n=n()) %>%
1919 mutate(prop=n/sum(n))
1920
1921 a11 %>%
1922 select(HHID, current_smoker, never_tried_smoking, ex_smoker) %>%
1923 filter(HHID == "1974")
1924
1925 # data number #1974 is incorrectly coded, adjustment is made for this data according to the answers
    from the questionnaires
1926 a11 <- a11 %>%
1927   mutate(smoke = ifelse(HHID == "1974", "Ex Smoker",
1928     ifelse(ex_smoker == 0 & current_smoker == 0, "Non Smoker",
1929       ifelse(ex_smoker == 1 & current_smoker == 0, "Ex Smoker",
1930         "Current Smoker"))))
1931
1932 a11$smoke <- factor(a11$smoke, levels= c("Non Smoker", "Ex Smoker", "Current Smoker"))
1933
1934 prop_frame(a11, "smoke")
1935
1936
1937 ## Physical Activity
1938 a11$active[a11$active == "K"] <- NA
1939 a11$active[a11$active == "R"] <- NA
1940
1941 a11$active <- factor(a11$active, labels=c("Not Active", "Active"))
1942
1943 prop_frame(a11, "active")
1944
1945
1946 ## Sedentary Lifestyle
1947 a11$sedentary[a11$sedentary == "K"] <- NA
1948 a11$sedentary[a11$sedentary == "R"] <- NA
1949
1950 a11$sedentary <- factor(a11$sedentary, labels=c("Not Sedentary", "Sedentary"))
1951
1952 prop_frame(a11, "sedentary")
1953
1954
1955 ## Difficulty Climbing Several Steps of Stairs

```

```

1956 summary(a11$A4_03)
1957
1958 a11$A4_03[a11$A4_03 == "."] <- NA
1959 a11$A4_03[a11$A4_03 == ".K"] <- NA
1960 a11$A4_03[a11$A4_03 == ".R"] <- NA
1961 a11$stair <- factor(a11$A4_03, labels=c("A Lot Difficult", "A Little Difficult", "No Difficulty"))
1962
1963 prop_frame(a11, "stair")
1964
1965
1966
1967 ##### NZHS 2011/12 – child #####
1968
1969 summary(HS11C$bmi) # 1037 Unobtainable, 415 Refused, 19 Outliers
1970
1971 HS11C$bmiscale <- as.numeric(as.character(HS11C$bmi))
1972
1973 HS11C$gender <- as.factor(HS11C$gender) #0: male, 1: female
1974
1975 HS11C <- HS11C %>%
1976   mutate(under =
1977     ifelse(gender == 0 & age == 2 & bmiscale <15.24, 0,
1978     ifelse(gender == 0 & age == 3 & bmiscale <14.83, 0,
1979     ifelse(gender == 0 & age == 4 & bmiscale <14.51, 0,
1980     ifelse(gender == 0 & age == 5 & bmiscale <14.26, 0,
1981     ifelse(gender == 0 & age == 6 & bmiscale <14.06, 0,
1982     ifelse(gender == 0 & age == 7 & bmiscale <14.00, 0,
1983     ifelse(gender == 0 & age == 8 & bmiscale <14.13, 0,
1984     ifelse(gender == 0 & age == 9 & bmiscale <14.36, 0,
1985     ifelse(gender == 0 & age == 10 & bmiscale <14.63, 0,
1986     ifelse(gender == 0 & age == 11 & bmiscale <14.96, 0,
1987     ifelse(gender == 0 & age == 12 & bmiscale <15.36, 0,
1988     ifelse(gender == 0 & age == 13 & bmiscale <15.84, 0,
1989     ifelse(gender == 0 & age == 14 & bmiscale <16.39, 0,
1990     ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
1991     ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
1992     ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
1993     ifelse(gender == 1 & age == 2 & bmiscale <14.96, 0,
1994     ifelse(gender == 1 & age == 3 & bmiscale <14.60, 0,
1995     ifelse(gender == 1 & age == 4 & bmiscale <14.30, 0,

```

```

1996   ifelse (gender == 1 & age == 5 & bmiscale <14.04, 0,
1997   ifelse (gender == 1 & age == 6 & bmiscale <13.85, 0,
1998   ifelse (gender == 1 & age == 7 & bmiscale <13.83, 0,
1999   ifelse (gender == 1 & age == 8 & bmiscale <14.00, 0,
2000   ifelse (gender == 1 & age == 9 & bmiscale <14.26, 0,
2001   ifelse (gender == 1 & age == 10 & bmiscale <14.58, 0,
2002   ifelse (gender == 1 & age == 11 & bmiscale <15.03, 0,
2003   ifelse (gender == 1 & age == 12 & bmiscale <15.59, 0,
2004   ifelse (gender == 1 & age == 13 & bmiscale <16.23, 0,
2005   ifelse (gender == 1 & age == 14 & bmiscale <16.86, 0,
2006   ifelse (gender == 1 & age == 15 & bmiscale <17.43, 0,
2007   ifelse (gender == 1 & age == 16 & bmiscale <17.90, 0,
2008   ifelse (gender == 1 & age == 17 & bmiscale <18.24, 0,
2009   ifelse (age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))
2010
2011 HS11C <- HS11C %>%
2012   mutate (average =
2013   ifelse (gender == 0 & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
2014   ifelse (gender == 0 & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
2015   ifelse (gender == 0 & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
2016   ifelse (gender == 0 & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
2017   ifelse (gender == 0 & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
2018   ifelse (gender == 0 & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
2019   ifelse (gender == 0 & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
2020   ifelse (gender == 0 & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
2021   ifelse (gender == 0 & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
2022   ifelse (gender == 0 & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
2023   ifelse (gender == 0 & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
2024   ifelse (gender == 0 & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
2025   ifelse (gender == 0 & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
2026   ifelse (gender == 0 & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
2027   ifelse (gender == 0 & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
2028   ifelse (gender == 0 & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
2029   ifelse (gender == 1 & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
2030   ifelse (gender == 1 & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
2031   ifelse (gender == 1 & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
2032   ifelse (gender == 1 & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
2033   ifelse (gender == 1 & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
2034   ifelse (gender == 1 & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
2035   ifelse (gender == 1 & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,

```

```

2036 ifelse(gender == 1 & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
2037 ifelse(gender == 1 & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
2038 ifelse(gender == 1 & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
2039 ifelse(gender == 1 & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
2040 ifelse(gender == 1 & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
2041 ifelse(gender == 1 & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
2042 ifelse(gender == 1 & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
2043 ifelse(gender == 1 & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
2044 ifelse(gender == 1 & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
2045 ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA))))))))))))))))))))))))))))))
2046
2047 HS11C <- HS11C %>%
2048 mutate(over=
2049 ifelse(gender == 0 & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
2050 ifelse(gender == 0 & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
2051 ifelse(gender == 0 & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
2052 ifelse(gender == 0 & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
2053 ifelse(gender == 0 & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
2054 ifelse(gender == 0 & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
2055 ifelse(gender == 0 & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
2056 ifelse(gender == 0 & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
2057 ifelse(gender == 0 & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
2058 ifelse(gender == 0 & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
2059 ifelse(gender == 0 & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
2060 ifelse(gender == 0 & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
2061 ifelse(gender == 0 & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
2062 ifelse(gender == 0 & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
2063 ifelse(gender == 0 & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
2064 ifelse(gender == 0 & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
2065 ifelse(gender == 1 & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
2066 ifelse(gender == 1 & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
2067 ifelse(gender == 1 & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
2068 ifelse(gender == 1 & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
2069 ifelse(gender == 1 & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
2070 ifelse(gender == 1 & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
2071 ifelse(gender == 1 & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
2072 ifelse(gender == 1 & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
2073 ifelse(gender == 1 & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
2074 ifelse(gender == 1 & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
2075 ifelse(gender == 1 & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,

```

```

2076 ifelse(gender == 1 & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
2077 ifelse(gender == 1 & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
2078 ifelse(gender == 1 & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
2079 ifelse(gender == 1 & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
2080 ifelse(gender == 1 & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
2081 ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))
2082
2083 HS11C <- HS11C %>%
2084 mutate(obes=
2085 ifelse(gender == 0 & age == 2 & bmiscale >=19.99, 3,
2086 ifelse(gender == 0 & age == 3 & bmiscale >=19.50, 3,
2087 ifelse(gender == 0 & age == 4 & bmiscale >=19.23, 3,
2088 ifelse(gender == 0 & age == 5 & bmiscale >=19.27, 3,
2089 ifelse(gender == 0 & age == 6 & bmiscale >=19.76, 3,
2090 ifelse(gender == 0 & age == 7 & bmiscale >=20.59, 3,
2091 ifelse(gender == 0 & age == 8 & bmiscale >=21.56, 3,
2092 ifelse(gender == 0 & age == 9 & bmiscale >=22.71, 3,
2093 ifelse(gender == 0 & age == 10 & bmiscale >=23.96, 3,
2094 ifelse(gender == 0 & age == 11 & bmiscale >=25.07, 3,
2095 ifelse(gender == 0 & age == 12 & bmiscale >=26.02, 3,
2096 ifelse(gender == 0 & age == 13 & bmiscale >=26.87, 3,
2097 ifelse(gender == 0 & age == 14 & bmiscale >=27.64, 3,
2098 ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
2099 ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
2100 ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
2101 ifelse(gender == 1 & age == 2 & bmiscale >=19.81, 3,
2102 ifelse(gender == 1 & age == 3 & bmiscale >=19.38, 3,
2103 ifelse(gender == 1 & age == 4 & bmiscale >=19.16, 3,
2104 ifelse(gender == 1 & age == 5 & bmiscale >=19.20, 3,
2105 ifelse(gender == 1 & age == 6 & bmiscale >=19.61, 3,
2106 ifelse(gender == 1 & age == 7 & bmiscale >=20.39, 3,
2107 ifelse(gender == 1 & age == 8 & bmiscale >=21.44, 3,
2108 ifelse(gender == 1 & age == 9 & bmiscale >=22.66, 3,
2109 ifelse(gender == 1 & age == 10 & bmiscale >=23.97, 3,
2110 ifelse(gender == 1 & age == 11 & bmiscale >=25.25, 3,
2111 ifelse(gender == 1 & age == 12 & bmiscale >=26.47, 3,
2112 ifelse(gender == 1 & age == 13 & bmiscale >=27.57, 3,
2113 ifelse(gender == 1 & age == 14 & bmiscale >=28.42, 3,
2114 ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
2115 ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,

```

```

2116   ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
2117   ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))
2118
2119
2120 # 0: underweight, 1: average, 2: overweight, 3: obese
2121 HS11C <- HS11C %>%
2122   mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
2123     ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
2124     ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
2125     ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA))))))
2126
2127 HS11C$bmic <- as.factor(HS11C$bmic)
2128
2129
2130 # excluded underweight in the bmi data
2131 c11 <- subset(HS11C, !bmic == 0 | is.na(bmic))
2132
2133
2134 ## BMI
2135 c11$bmic <- factor(c11$bmic, labels=c("Average", "Overweight", "Obese"))
2136
2137 prop_frame(c11, "bmic")
2138
2139
2140 ## Sex
2141 c11$gender <- factor(c11$gender, labels=c("Male", "Female"))
2142
2143 prop_frame(c11, "gender")
2144
2145
2146 ## Age
2147 summary(c11$age)
2148
2149
2150 ## Ethnicity
2151 # recode anyone ~ european
2152 c11 = c11 %>%
2153   mutate(euro = ifelse(C4_03_01 == 1, 1, 0))
2154
2155 # recode anyone ~ maori

```

```

2156 c11 = c11 %>%
2157   mutate(maori = ifelse(C4_03_02 == 1, 1, 0))
2158
2159 # recode anyone ~ pacific
2160 c11 = c11 %>%
2161   mutate(pacific = ifelse(C4_03_03 == 1, 1,
2162     ifelse(C4_03_04 == 1, 1,
2163       ifelse(C4_03_05 == 1, 1,
2164         ifelse(C4_03_06 == 1, 1, 0))))))
2165
2166 # recode anyone ~ asian (Indian and Chinese)
2167 c11 = c11 %>%
2168   mutate(asian = ifelse(C4_03_07 == 1, 1,
2169     ifelse(C4_03_08 == 1, 1, 0)))
2170
2171 # recode anyone ~ other
2172 c11 = c11 %>%
2173   mutate(other = ifelse(C4_03_77 == 1, 1, 0))
2174
2175
2176 c11 = c11 %>%
2177   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
2178     Only",
2179     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
2180       Pacific Only",
2181       ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
2182         Only",
2183         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
2184           European Only",
2185           ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
2186             ,
2187             ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
2188               Ethnicities (M)",
2189               ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
2190                 Ethnicities (M)",
2191                 ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
2192                   ,
2193                   ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
2194                     ,

```

```

ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
,
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
,
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
,
ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
,
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
,
ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
,
ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
,
ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
,
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
,

```



```

2206         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
           Ethnicities (M)",
2207         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
           Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))
2208
2209 c11$eth_count <- as.factor(c11$eth_count)
2210
2211 prop_frame(c11, "eth_count")
2212
2213
2214 ## Household Income
2215 # merge adult 2011 with child 2011 data by HHID to match income and education data from the same
      household
2216 ac11 <- left_join(select(c11, HHID), select(a11, HHID, A5_24, A5_14, A5_15), by="HHID")
2217
2218 summary(ac11$A5_24)
2219
2220 ac11$hhinc <- as.numeric(as.character(ac11$A5_24))
2221 ac11$hhinc <- cut(ac11$hhinc, breaks=c(0,5,10,13,16))
2222 ac11$hhinc <- factor(ac11$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000", ">$70,000"
      ))
2223
2224 # return the value to the original dataset
2225 c11$hhinc <- ac11$hhinc
2226
2227 prop_frame(c11, "hhinc")
2228
2229
2230 ## Education
2231 ac11$secondary <- as.numeric(as.character(ac11$A5_14))
2232 ac11$secondary <- recode(ac11$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
      school grads
2233 ac11$secondary <- as.factor(ac11$secondary)
2234
2235 ac11$tertiary <- as.numeric(as.character(ac11$A5_15))
2236 ac11$tertiary <- recode(ac11$tertiary, "0" = "0", "1"="2",
      "2"="2", "3"="2", "4"="2", .default = "3") #0 = no tertiary, 2= secondary, 3=
      tertiary
2237
2238 ac11$tertiary <- as.factor(ac11$tertiary)
2239

```

```

2240 # prioritise the highest educational qualification
2241 ac11 <- mutate(ac11, edu = ifelse(secondary == "1" & tertiary == "0", 1,
2242                                   ifelse(secondary == "2" & tertiary == "0", 2,
2243                                           ifelse(tertiary == "2", 2,
2244                                                   ifelse(tertiary == "3", 3, 1))))))
2245
2246 # return the information to original data set
2247 c11$edu <- ac11$edu
2248 c11$edu <- as.factor(c11$edu)
2249 c11$edu <- factor(c11$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
2250
2251 prop_frame(c11, "edu")
2252
2253 # remove ac11 from the memory
2254 rm(ac11)
2255
2256 # education missing value
2257 sum(is.na(c11$edu))
2258 sum(is.na(c11$edu))/length(c11$edu) # 455 NA (15.27%)
2259
2260
2261 ## Deprivation Quintile
2262 names(c11)[names(c11) == "nzdep06_quintile"] <- "dep"
2263
2264 prop_frame(c11, "dep")
2265
2266
2267 ## Fruit and Vegetable
2268 c11$veges <- c11$kidveges_new
2269 c11$fruit <- c11$kidfruit_2serves
2270
2271 c11$fruit[c11$fruit == "K"] <- NA
2272 c11$fruit[c11$fruit == "X"] <- NA
2273 c11$veges[c11$veges == "K"] <- NA
2274 c11$veges[c11$veges == "R"] <- NA
2275 c11$veges[c11$veges == "X"] <- NA
2276 c11$fruit <- factor(c11$fruit, labels=c("No", "Yes"))
2277 c11$veges <- factor(c11$veges, labels=c("No", "Yes"))
2278
2279 prop_frame(c11, "fruit")

```

```

2280 prop_frame(c11, "veges")
2281
2282
2283 ## Soft Drink and Fast Food
2284 summary(c11$C3_09) #14 NA, 18 K
2285 c11$C3_09 <- as.numeric(as.character(c11$C3_09))
2286
2287 c11$softd <- cut(c11$C3_09, c(-Inf,1,2,4,Inf), right=FALSE)
2288 summary(c11$softd)
2289
2290 c11$softd <- factor(c11$softd, labels=c("0/week", "1/week", "2-3/week", "4+/week"))
2291
2292 prop_frame(c11, "softd")
2293
2294 #fast food
2295 summary(c11$C3_10) # 14 NA, 8 K
2296 c11$C3_10 <- as.numeric(as.character(c11$C3_10))
2297
2298 c11$fastf <- cut(c11$C3_10, c(0,1,2,4,Inf), right=FALSE)
2299
2300 c11$fastf <- factor(c11$fastf, labels=c("0/week", "1/week", "2-3/week", "4+/week"))
2301
2302 prop_frame(c11, "fastf")
2303
2304
2305
2306 ##### NZHS 2012/13 – adult #####
2307
2308 summary(HS12A$bmi) # 427 R, 423 U, 214 P, 4 O
2309
2310 HS12A$bmiscale <- as.numeric(as.character(HS12A$bmi)) # force non-numeric into NA
2311
2312 HS12A$gender <- as.factor(HS12A$gender)
2313
2314 HS12A <- HS12A %>%
2315   mutate(bmic = ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
2316     ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
2317       ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
2318         ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
2319           ifelse(gender == 1 & age == 16 & bmiscale <17.9, 0,

```

```

2320     ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
2321     ifelse(age >= 18 & bmiscale <18.5, 0,
2322     ifelse(gender == 0 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
2323     ifelse(gender == 0 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
2324     ifelse(gender == 0 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
2325     ifelse(gender == 1 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
2326     ifelse(gender == 1 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
2327     ifelse(gender == 1 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
2328     ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
2329     ifelse(gender == 0 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
2330     ifelse(gender == 0 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
2331     ifelse(gender == 0 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
2332     ifelse(gender == 1 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
2333     ifelse(gender == 1 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
2334     ifelse(gender == 1 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
2335     ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
2336     ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
2337     ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
2338     ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
2339     ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
2340     ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
2341     ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
2342     ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))
2343
2344 HS12A$bmic <- as.factor(HS12A$bmic)
2345
2346 # exclude underweight
2347 a12 <- subset(HS12A, !bmic == 0 | is.na(bmic))
2348
2349
2350 ## Migration Status
2351 # recode people who lived in NZ for less than 10–11 years as Migrants
2352 a12$native <- recode(a12$A5_06, "2003"="Migrant", "2004"="Migrant", "2005"="Migrant",
2353     "2006"="Migrant", "2007"="Migrant", "2008"="Migrant", "2009"="
2354     Migrant",
2355     "2010"="Migrant", "2011"="Migrant", "2012"="Migrant", ".K"= "S", ".R
2356     "= "S",
2357     .default="Native")
2358 a12$native[a12$native == "S"] <- NA #recode S into NA
2359 a12$native <- droplevels(a12$native)

```

```

2358
2359 prop_frame(a12, "native")
2360
2361
2362 ## BMI
2363 a12$bmic <- factor(a12$bmic, labels =c("Average", "Overweight", "Obese"))
2364
2365 prop_frame(a12, "bmic")
2366
2367
2368 ## Sex
2369 a12$gender <- factor(a12$gender, labels=c("Male", "Female"))
2370
2371 prop_frame(a12, "gender")
2372
2373
2374 ## Age
2375 summary(a12$age)
2376
2377
2378 ## Ethnicity
2379 # recode anyone ~ european
2380 a12 = a12 %>%
2381   mutate(euro = ifelse(A5_03_01 == 1, 1, 0))
2382
2383 # recode anyone ~ maori
2384 a12 = a12 %>%
2385   mutate(maori = ifelse(A5_03_02 == 1, 1, 0))
2386
2387 # recode anyone ~ pacific
2388 a12 = a12 %>%
2389   mutate(pacific = ifelse(A5_03_03 == 1, 1,
2390     ifelse(A5_03_04 == 1, 1,
2391       ifelse(A5_03_05 == 1, 1,
2392         ifelse(A5_03_06 == 1, 1, 0))))))
2393
2394 # recode anyone ~ asian
2395 a12 = a12 %>%
2396   mutate(asian = ifelse(A5_03_07 == 1, 1,
2397     ifelse(A5_03_08 == 1, 1, 0)))

```

```

2398
2399 # recode anyone ~ other
2400 a12 = a12 %>%
2401   mutate(other = ifelse(A5_03_77 == 1, 1, 0))
2402
2403
2404 a12 = a12 %>%
2405   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
2406     Only",
2407     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
2408       Pacific Only",
2409       ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
2410         Only",
2411         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
2412           European Only",
2413           ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
2414             ,
2415             ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
2416               Ethnicities (M)",
2417               ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
2418                 Ethnicities (M)",
2419                 ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
2420                   ,
2421                   ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
2422                     ,
2423                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
2424                       ,
2425                       ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
2426                         Ethnicities (M)",
2427                         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
2428                           Ethnicities (M)",
2429                         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
2430                           ,
2431                           ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
2432                             ,
2433                             ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
2434                               ,
2435                               ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
2436                                 Ethnicities (M)",

```

```

2421         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
2422             Ethnicities (M)",
2423         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
2424             Ethnicities (M)",
2425         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
2426             ,
2427         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
2428             ,
2429         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
2430             ,
2431         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
2432             Ethnicities (M)",
2433         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
2434             Ethnicities (M)",
2435         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
2436             Ethnicities (M)",
2437         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
2438             ,
2439         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
2440             Ethnicities (M)",
2441         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
2442             Ethnicities (M)",
2443         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
2444             Ethnicities (M)",
2445         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
2446             ,
2447         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
2448             Ethnicities (M)",
2449         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
2450             Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))))))
2451
2452 a12$eth_count <- as.factor(a12$eth_count)
2453
2454 prop_frame(a12, "eth_count")
2455
2456
2457 ## Household Income
2458 a12$hhinc <- ordered(a12$A5_24, levels=c("4", "5", "6", "7", "8", "9", "10", "11", "12",
2459                                           "13", "14", "15", "16"))
2460

```

```

2446 a12$hhinc <- as.numeric(as.character(a12$hhinc))
2447 a12$hhinc <- cut(a12$hhinc, breaks=c(0,5,10,13,16))
2448 a12$hhinc <- factor(a12$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000", ">$70,000"))
2449
2450 prop_frame(a12, "hhinc")
2451
2452
2453 ## Education
2454 a12$secondary <- as.numeric(as.character(a12$A5_14))
2455 a12$secondary <- recode(a12$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
2456 a12$secondary <- as.factor(a12$secondary)
2457
2458 a12$tertiary <- as.numeric(as.character(a12$A5_15)) # force .K and R. into NA
2459 a12$tertiary <- recode(a12$tertiary, "0" = "0", "1"="2", "2"="2",
    "3"="2", "4"="2", .default = "3") #0 = no tertiary, 2= secondary, 3= tertiary
2460
2461 a12$tertiary <- as.factor(a12$tertiary)
2462
2463 a12 <- mutate(a12, edu = ifelse(secondary == "1" & tertiary == "0", 1,
    ifelse(secondary == "2" & tertiary == "0", 2,
    ifelse(tertiary == "2", 2,
    ifelse(tertiary == "3", 3, 1))))
2464
2465
2466
2467
2468 a12$edu <- as.factor(a12$edu)
2469
2470 a12$edu <- factor(a12$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
2471
2472 prop_frame(a12, "edu")
2473
2474
2475 ## Deprivation Quintile
2476 names(a12)[names(a12) == "nzdep06_quintile"] <- "dep"
2477
2478 prop_frame(a12, "dep")
2479
2480
2481 ## Fruit and Vegetable
2482 summary(a12$fruit) # 10 K, 1 R
2483 a12$fruit[a12$fruit == "K"] <- NA
2484 a12$fruit[a12$fruit == "R"] <- NA

```



```

2485
2486 a12$fruit <- factor(a12$fruit, labels=c("No", "Yes"))
2487
2488 prop_frame(a12, "fruit")
2489
2490 summary(a12$veges) # 7 K, 1 R
2491 a12$veges[a12$veges == "K"] <- NA
2492 a12$veges[a12$veges == "R"] <- NA
2493
2494 a12$veges <- factor(a12$veges, labels=c("No", "Yes"))
2495
2496 prop_frame(a12, "veges")
2497
2498
2499 ## Alcohol Problom
2500 summary(a12$haz_drinker_all) # 130 K, 14 R
2501
2502 a12$haz_drinker_all[a12$haz_drinker_all == "K"] <- NA
2503 a12$haz_drinker_all[a12$haz_drinker_all == "R"] <- NA
2504
2505 a12$haz_drinker_all <- factor(a12$haz_drinker_all, labels=c("No Alcohol Problem",
2506                                                         "Alcohol Problem"))
2507
2508 prop_frame(a12, "haz_drinker_all")
2509
2510
2511 ## Smoking Status
2512 a12 <- a12 %>%
2513   mutate(smoke = ifelse(current_smoker == 1 & ex_smoker == 0, 3,
2514                         ifelse(current_smoker == 0 & ex_smoker == 1, 2,
2515                               ifelse(current_smoker == "K", NA,
2516                                     ifelse(current_smoker == "R", NA,
2517                                             ifelse(ex_smoker == "K", NA,
2518                                                   ifelse(ex_smoker == "R", NA, 1)))))))
2519
2520 a12$smoke <- as.factor(a12$smoke)
2521 a12$smoke <- factor(a12$smoke, labels=c("Non Smoker", "Ex Smoker", "Current Smoker"))
2522
2523 prop_frame(a12, "smoke")
2524

```

```

2525
2526 ## Physically Active
2527 summary(a12$active) # 131 K, 3 R
2528 a12$active[a12$active == "K"] <- NA
2529 a12$active[a12$active == "R"] <- NA
2530
2531 a12$active <- factor(a12$active, labels=c("Not Active", "Active"))
2532
2533 summary(a12$sedentary) # 129 K, 2 R
2534
2535 a12$sedentary[a12$sedentary == "K"] <- NA
2536 a12$sedentary[a12$sedentary == "R"] <- NA
2537
2538 a12$sedentary <- factor(a12$sedentary, labels=c("Not Sedentary", "Sedentary"))
2539
2540 prop_frame(a12, "active")
2541 prop_frame(a12, "sedentary")
2542
2543
2544 ## Difficulty Climbing Several Steps of Stairs
2545 summary(a12$A4_03)
2546 a12$A4_03[a12$A4_03 == ".K"] <- NA
2547 a12$A4_03[a12$A4_03 == ".R"] <- NA
2548
2549 a12$stair <- factor(a12$A4_03, labels=c("A Lot Difficult", "A Little Difficult",
2550                                         "No Difficulty"))
2551
2552 prop_frame(a12, "stair")
2553
2554
2555
2556 ##### NZHS 2012/13 – child #####
2557
2558 ### SUBSETTING ###
2559 HS12C$bmiscale <- as.numeric(as.character(HS12C$bmi))
2560
2561 HS12C$gender <- as.factor(HS12C$gender)
2562
2563 HS12C <- HS12C %>%
2564   mutate(under =

```

```

2565   ifelse (gender == 0 & age == 2 & bmiscale <15.24, 0,
2566   ifelse (gender == 0 & age == 3 & bmiscale <14.83, 0,
2567   ifelse (gender == 0 & age == 4 & bmiscale <14.51, 0,
2568   ifelse (gender == 0 & age == 5 & bmiscale <14.26, 0,
2569   ifelse (gender == 0 & age == 6 & bmiscale <14.06, 0,
2570   ifelse (gender == 0 & age == 7 & bmiscale <14.00, 0,
2571   ifelse (gender == 0 & age == 8 & bmiscale <14.13, 0,
2572   ifelse (gender == 0 & age == 9 & bmiscale <14.36, 0,
2573   ifelse (gender == 0 & age == 10 & bmiscale <14.63, 0,
2574   ifelse (gender == 0 & age == 11 & bmiscale <14.96, 0,
2575   ifelse (gender == 0 & age == 12 & bmiscale <15.36, 0,
2576   ifelse (gender == 0 & age == 13 & bmiscale <15.84, 0,
2577   ifelse (gender == 0 & age == 14 & bmiscale <16.39, 0,
2578   ifelse (gender == 0 & age == 15 & bmiscale <16.98, 0,
2579   ifelse (gender == 0 & age == 16 & bmiscale <17.53, 0,
2580   ifelse (gender == 0 & age == 17 & bmiscale <18.04, 0,
2581   ifelse (gender == 1 & age == 2 & bmiscale <14.96, 0,
2582   ifelse (gender == 1 & age == 3 & bmiscale <14.60, 0,
2583   ifelse (gender == 1 & age == 4 & bmiscale <14.30, 0,
2584   ifelse (gender == 1 & age == 5 & bmiscale <14.04, 0,
2585   ifelse (gender == 1 & age == 6 & bmiscale <13.85, 0,
2586   ifelse (gender == 1 & age == 7 & bmiscale <13.83, 0,
2587   ifelse (gender == 1 & age == 8 & bmiscale <14.00, 0,
2588   ifelse (gender == 1 & age == 9 & bmiscale <14.26, 0,
2589   ifelse (gender == 1 & age == 10 & bmiscale <14.58, 0,
2590   ifelse (gender == 1 & age == 11 & bmiscale <15.03, 0,
2591   ifelse (gender == 1 & age == 12 & bmiscale <15.59, 0,
2592   ifelse (gender == 1 & age == 13 & bmiscale <16.23, 0,
2593   ifelse (gender == 1 & age == 14 & bmiscale <16.86, 0,
2594   ifelse (gender == 1 & age == 15 & bmiscale <17.43, 0,
2595   ifelse (gender == 1 & age == 16 & bmiscale <17.90, 0,
2596   ifelse (gender == 1 & age == 17 & bmiscale <18.24, 0,
2597   ifelse (age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))
2598
2599 HS12C <- HS12C %>%
2600   mutate (average =
2601     ifelse (gender == 0 & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
2602     ifelse (gender == 0 & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
2603     ifelse (gender == 0 & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
2604     ifelse (gender == 0 & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,

```

```

2605 ifelse(gender == 0 & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
2606 ifelse(gender == 0 & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
2607 ifelse(gender == 0 & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
2608 ifelse(gender == 0 & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
2609 ifelse(gender == 0 & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
2610 ifelse(gender == 0 & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
2611 ifelse(gender == 0 & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
2612 ifelse(gender == 0 & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
2613 ifelse(gender == 0 & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
2614 ifelse(gender == 0 & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
2615 ifelse(gender == 0 & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
2616 ifelse(gender == 0 & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
2617 ifelse(gender == 1 & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
2618 ifelse(gender == 1 & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
2619 ifelse(gender == 1 & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
2620 ifelse(gender == 1 & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
2621 ifelse(gender == 1 & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
2622 ifelse(gender == 1 & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
2623 ifelse(gender == 1 & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
2624 ifelse(gender == 1 & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
2625 ifelse(gender == 1 & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
2626 ifelse(gender == 1 & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
2627 ifelse(gender == 1 & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
2628 ifelse(gender == 1 & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
2629 ifelse(gender == 1 & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
2630 ifelse(gender == 1 & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
2631 ifelse(gender == 1 & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
2632 ifelse(gender == 1 & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
2633 ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA))))))))))))))))))))))))))))))
2634
2635 HS12C <- HS12C %>%
2636 mutate(over=
2637 ifelse(gender == 0 & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
2638 ifelse(gender == 0 & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
2639 ifelse(gender == 0 & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
2640 ifelse(gender == 0 & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
2641 ifelse(gender == 0 & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
2642 ifelse(gender == 0 & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
2643 ifelse(gender == 0 & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
2644 ifelse(gender == 0 & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,

```

```

2645 ifelse(gender == 0 & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
2646 ifelse(gender == 0 & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
2647 ifelse(gender == 0 & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
2648 ifelse(gender == 0 & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
2649 ifelse(gender == 0 & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
2650 ifelse(gender == 0 & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
2651 ifelse(gender == 0 & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
2652 ifelse(gender == 0 & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
2653 ifelse(gender == 1 & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
2654 ifelse(gender == 1 & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
2655 ifelse(gender == 1 & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
2656 ifelse(gender == 1 & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
2657 ifelse(gender == 1 & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
2658 ifelse(gender == 1 & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
2659 ifelse(gender == 1 & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
2660 ifelse(gender == 1 & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
2661 ifelse(gender == 1 & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
2662 ifelse(gender == 1 & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
2663 ifelse(gender == 1 & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
2664 ifelse(gender == 1 & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
2665 ifelse(gender == 1 & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
2666 ifelse(gender == 1 & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
2667 ifelse(gender == 1 & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
2668 ifelse(gender == 1 & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
2669 ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))
2670
2671 HS12C <- HS12C %>%
2672 mutate(obes=
2673 ifelse(gender == 0 & age == 2 & bmiscale >=19.99, 3,
2674 ifelse(gender == 0 & age == 3 & bmiscale >=19.50, 3,
2675 ifelse(gender == 0 & age == 4 & bmiscale >=19.23, 3,
2676 ifelse(gender == 0 & age == 5 & bmiscale >=19.27, 3,
2677 ifelse(gender == 0 & age == 6 & bmiscale >=19.76, 3,
2678 ifelse(gender == 0 & age == 7 & bmiscale >=20.59, 3,
2679 ifelse(gender == 0 & age == 8 & bmiscale >=21.56, 3,
2680 ifelse(gender == 0 & age == 9 & bmiscale >=22.71, 3,
2681 ifelse(gender == 0 & age == 10 & bmiscale >=23.96, 3,
2682 ifelse(gender == 0 & age == 11 & bmiscale >=25.07, 3,
2683 ifelse(gender == 0 & age == 12 & bmiscale >=26.02, 3,
2684 ifelse(gender == 0 & age == 13 & bmiscale >=26.87, 3,

```

```

2685 ifelse(gender == 0 & age == 14 & bmiscale >=27.64, 3,
2686 ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
2687 ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
2688 ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
2689 ifelse(gender == 1 & age == 2 & bmiscale >=19.81, 3,
2690 ifelse(gender == 1 & age == 3 & bmiscale >=19.38, 3,
2691 ifelse(gender == 1 & age == 4 & bmiscale >=19.16, 3,
2692 ifelse(gender == 1 & age == 5 & bmiscale >=19.20, 3,
2693 ifelse(gender == 1 & age == 6 & bmiscale >=19.61, 3,
2694 ifelse(gender == 1 & age == 7 & bmiscale >=20.39, 3,
2695 ifelse(gender == 1 & age == 8 & bmiscale >=21.44, 3,
2696 ifelse(gender == 1 & age == 9 & bmiscale >=22.66, 3,
2697 ifelse(gender == 1 & age == 10 & bmiscale >=23.97, 3,
2698 ifelse(gender == 1 & age == 11 & bmiscale >=25.25, 3,
2699 ifelse(gender == 1 & age == 12 & bmiscale >=26.47, 3,
2700 ifelse(gender == 1 & age == 13 & bmiscale >=27.57, 3,
2701 ifelse(gender == 1 & age == 14 & bmiscale >=28.42, 3,
2702 ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
2703 ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
2704 ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
2705 ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))
2706
2707 # 0: underweight, 1: average, 2: overweight, 3: obese
2708 HS12C <- HS12C %>%
2709   mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
2710     ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
2711     ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
2712     ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA))))
2713
2714 HS12C$bmic <- as.factor(HS12C$bmic)
2715
2716 # exclude underweight
2717 c12 <- subset(HS12C, !bmic == 0 | is.na(bmic))
2718
2719
2720 ## BMI
2721 c12$bmic <- factor(c12$bmic, labels=c("Average", "Overweight", "Obese"))
2722
2723 prop_frame(c12, "bmic")
2724

```

```

2725
2726 ## Sex
2727 c12$gender <- factor(c12$gender, labels=c("Male", "Female"))
2728
2729 prop_frame(c12, "gender")
2730
2731
2732 ## Age
2733 summary(c12$age)
2734
2735
2736 ## Ethnicity
2737 # recode anyone ~ european
2738 c12 = c12 %>%
2739   mutate(euro = ifelse(C4_03_01 == 1, 1, 0))
2740
2741 # recode anyone ~ maori
2742 c12 = c12 %>%
2743   mutate(maori = ifelse(C4_03_02 == 1, 1, 0))
2744
2745 # recode anyone ~ pacific
2746 c12 = c12 %>%
2747   mutate(pacific = ifelse(C4_03_03 == 1, 1,
2748     ifelse(C4_03_04 == 1, 1,
2749       ifelse(C4_03_05 == 1, 1,
2750         ifelse(C4_03_06 == 1, 1, 0))))))
2751
2752 # recode anyone ~ asian
2753 # starting from 2011, other asian will be coded as other ethnicity because there is no way to
2754   identify other asian
2755 c12 = c12 %>%
2756   mutate(asian = ifelse(C4_03_07 == 1, 1,
2757     ifelse(C4_03_08 == 1, 1, 0)))
2758
2759 # recode anyone ~ other
2760 c12 = c12 %>%
2761   mutate(other = ifelse(C4_03_77 == 1, 1, 0))
2762
2763 c12 = c12 %>%

```

```

2764 mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
    Only",
2765
    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
        Pacific Only",
2766
    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
        Only",
2767
    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
        European Only",
2768
    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
        ,
2769
    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
        Ethnicities (M)",
2770
    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
        Ethnicities (M)",
2771
    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
        ,
2772
    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
        ,
2773
    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
        ,
2774
    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
        Ethnicities (M)",
2775
    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
        Ethnicities (M)",
2776
    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
        ,
2777
    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
        ,
2778
    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
        ,
2779
    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
        Ethnicities (M)",
2780
    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
        Ethnicities (M)",
2781
    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
        Ethnicities (M)",
2782
    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
        ,
2783
    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
        ,

```



```

2784         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
2785         ,
2786         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
2787             Ethnicities (M)",
2788         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
2789             Ethnicities (M)",
2790         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
2791             Ethnicities (M)",
2792         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
2793         ,
2794         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
2795             Ethnicities (M)",
2796         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
2797             Ethnicities (M)",
2798         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
2799             Ethnicities (M)",
2800         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
2801         ,
2802         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
2803             Ethnicities (M)",
2804         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
2805             Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))))))
2806
2807 c12$eth_count <- as.factor(c12$eth_count)
2808
2809 prop_frame(c12, "eth_count")
2810
2811
2812 ## Household Income
2813 # merge adult data to child data set
2814 ac12 <- left_join(select(c12, HHID, C4_17), select(a12, HHID, A5_24, A5_14, A5_15), by="HHID")
2815
2816 ac12$ainc <- as.numeric(as.character(ac12$A5_24))
2817 ac12$cinc <- as.numeric(as.character(ac12$C4_17))
2818
2819 ac12 <- mutate(ac12, inc= ifelse(is.na(ainc), cinc,
2820                                ifelse(is.na(cinc), ainc, cinc)))
2821
2822 ac12$hhinc <- cut(ac12$inc, breaks=c(0,5,10,13,16))

```

```

2812 ac12$hhinc <- factor(ac12$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000", ">$70,000"
      ))
2813
2814 # copy it back to original data set
2815 c12$hhinc <- ac12$hhinc
2816
2817 prop_frame(c12, "hhinc")
2818
2819
2820 ## Education
2821 ac12$secondary <- as.numeric(as.character(ac12$A5_14))
2822 ac12$secondary <- recode(ac12$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
      school grads
2823 ac12$secondary <- as.factor(ac12$secondary)
2824
2825 ac12$tertiary <- as.numeric(as.character(ac12$A5_15))
2826 ac12$tertiary <- recode(ac12$tertiary, "0" = "0", "1" = "2", "2" = "2",
      "3" = "2", "4" = "2", .default = "3") #0 = no tertiary, 3= tertiary
2827
2828 ac12$tertiary <- as.factor(ac12$tertiary)
2829
2830 # prioritise highest education for duplicates
2831 ac12 <- mutate(ac12, edu = ifelse(secondary == "1" & tertiary == "0", 1,
      ifelse(secondary == "2" & tertiary == "0", 2,
      ifelse(tertiary == "2", 2,
      ifelse(tertiary == "3", 3, 1))))))
2832
2833
2834
2835
2836 ac12$edu <- as.factor(ac12$edu)
2837
2838 ac12$edu <- factor(ac12$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
2839
2840 # move them to the original data set
2841 c12$edu <- ac12$edu
2842
2843 prop_frame(c12, "edu")
2844
2845 # remove ac12 from the memory
2846 rm(ac12)
2847
2848
2849 ## Deprivation Quintile

```

```

2850 names(c12)[names(c12) == "nzdep06_quintile"] <- "dep"
2851
2852 prop_frame(c12, "dep")
2853
2854
2855 ## Fruit and Vegetable
2856 summary(c12$kidfruit_2serves) # 5 K
2857 summary(c12$kidveges_new) # 2K, 1R
2858 c12$kidfruit_2serves[c12$kidfruit_2serves == "K"] <- NA
2859 c12$kidfruit_2serves[c12$kidfruit_2serves == "X"] <- NA
2860 c12$kidveges_new[c12$kidveges_new == "K"] <- NA
2861 c12$kidveges_new[c12$kidveges_new == "R"] <- NA
2862 c12$kidveges_new[c12$kidveges_new == "X"] <- NA
2863
2864 c12$fruit <- factor(c12$kidfruit_2serves, labels=c("No", "Yes"))
2865 c12$veges <- factor(c12$kidveges_new, labels=c("No", "Yes"))
2866
2867 prop_frame(c12, "fruit")
2868 prop_frame(c12, "veges")
2869
2870
2871 ## Soft Drink and Fast Food
2872 summary(c12$C3_09) # 1 NA, 19 K, 1 R
2873 c12$softd <- as.numeric(as.character(c12$C3_09))
2874
2875 c12$softd <- cut(c12$softd, c(0,1,2,4,Inf), labels=c("0/week",
2876 "1/week", "2-3/week", "4+/week"), right=FALSE)
2877 summary(c12$softd)
2878
2879
2880 summary(c12$C3_10) # 1 NA, 8 K, 1 R
2881 c12$fastf <- as.numeric(as.character(c12$C3_10))
2882
2883 c12$fastf <- cut(c12$fastf, c(0,1,2,4,Inf), labels=c("0/week", "1/week",
2884 "2-3/week", "4+/week"), right=FALSE)
2885 summary(c12$fastf)
2886
2887 prop_frame(c12, "softd")
2888 prop_frame(c12, "fastf")
2889

```

```

2890
2891
2892 ##### NZHS 2013/14 – adult #####
2893
2894 summary(HS13A$bmi) # 261 R, 260 U, 194 P
2895
2896 HS13A$gender <- as.factor(HS13A$gender)
2897
2898 HS13A$bmiscale <- as.numeric(as.character(HS13A$bmi))
2899
2900 # Recode BMI category according to IOTF cut-offs
2901 HS13A <- HS13A %>%
2902   mutate(bmic = ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
2903     ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
2904       ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
2905         ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
2906           ifelse(gender == 1 & age == 16 & bmiscale <17.9, 0,
2907             ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
2908               ifelse(age >= 18 & bmiscale <18.5, 0,
2909                 ifelse(gender == 0 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
2910                   ifelse(gender == 0 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
2911                     ifelse(gender == 0 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
2912                       ifelse(gender == 1 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
2913                         ifelse(gender == 1 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
2914                           ifelse(gender == 1 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
2915                             ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
2916                               ifelse(gender == 0 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
2917                                 ifelse(gender == 0 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
2918                                   ifelse(gender == 0 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
2919                                     ifelse(gender == 1 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
2920                                       ifelse(gender == 1 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
2921                                         ifelse(gender == 1 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
2922                                           ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
2923                                             ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
2924                                               ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
2925                                                 ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
2926                                                   ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
2927                                                     ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
2928                                                       ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
2929                                                         ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))

```

```

2930
2931 HS13A$bmic <- as.factor(HS13A$bmic)
2932
2933 # exclude Underweight
2934 a13 <- subset(HS13A, !bmic == 0 | is.na(bmic))
2935
2936
2937 ## Migration Status
2938 a13$native <- recode(a13$A5_06, "2004"="Migrant", "2005"="Migrant", "2006"="Migrant", "2007"="
      Migrant",
2939                                "2008"="Migrant", "2009"="Migrant", "2010"="Migrant", "2011"="
      Migrant",
2940                                "2012"="Migrant", "2013"="Migrant", "2014"="Migrant", ".K"= "S", ".R
      "= "S",
2941                                .default="Native")
2942
2943 a13$native[a13$native == "S"] <- NA #recode S into NA
2944 a13$native <- droplevels(a13$native)
2945
2946 prop_frame(a13, "native")
2947
2948
2949 ## BMI
2950 a13$bmic <- factor(a13$bmic, labels=c("Average", "Overweight", "Obese"))
2951
2952 prop_frame(a13, "bmic")
2953
2954
2955 ### Sex
2956 a13$gender <- factor(a13$gender, labels=c("Male", "Female"))
2957
2958 prop_frame(a13, "gender")
2959
2960
2961 ### Age
2962 summary(a13$age)
2963
2964
2965 ## Ethnicity
2966 # recode anyone ~ european

```

```

2967 a13 = a13 %>%
2968   mutate(euro = ifelse(A5_03_01 == 1, 1, 0))
2969
2970 # recode anyone ~ maori
2971 a13 = a13 %>%
2972   mutate(maori = ifelse(A5_03_02 == 1, 1, 0))
2973
2974 # recode anyone ~ pacific
2975 a13 = a13 %>%
2976   mutate(pacific = ifelse(A5_03_03 == 1, 1,
2977     ifelse(A5_03_04 == 1, 1,
2978       ifelse(A5_03_05 == 1, 1,
2979         ifelse(A5_03_06 == 1, 1, 0))))))
2980
2981 # recode anyone ~ asian (Indian and Chinese)
2982 a13 = a13 %>%
2983   mutate(asian = ifelse(A5_03_07 == 1, 1,
2984     ifelse(A5_03_08 == 1, 1, 0)))
2985
2986 # recode anyone ~ other
2987 a13 = a13 %>%
2988   mutate(other = ifelse(A5_03_77 == 1, 1, 0))
2989
2990
2991 a13 = a13 %>%
2992   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
2993     Only",
2994       ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
2995         Pacific Only",
2996         ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
2997           Only",
2998           ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
2999             European Only",
3000             ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
3001               ,
3002               ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
3003                 Ethnicities (M)",
3004                 ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
3005                   Ethnicities (M)",

```

```

ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
,
ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
,
ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
,
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
,
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
,
ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
,
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
,
ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
,
ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
,
ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
Ethnicities (M)",
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
,
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
Ethnicities (M)",

```

```

3019         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
Ethnicities (M)",
3020         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
,
3021         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
Ethnicities (M)",
3022         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))
3023
3024 a13$eth_count <- as.factor(a13$eth_count)
3025
3026 prop_frame(a13, "eth_count")
3027
3028
3029 ## Household Income
3030 summary(a13$A5_24)
3031 a13$hhinc <- ordered(a13$A5_24, levels=c("4", "5", "6", "7", "8", "9", "10", "11", "12",
3032                                           "13", "14", "15", "16"))
3033
3034 a13$hhinc <- as.numeric(as.character(a13$hhinc))
3035 a13$hhinc <- cut(a13$hhinc, breaks=c(0,5,10,13,16))
3036 a13$hhinc <- factor(a13$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000", ">$70,000"))
3037
3038 prop_frame(a13, "hhinc")
3039
3040
3041 ## Education
3042 a13$secondary <- as.numeric(as.character(a13$A5_14))
3043 a13$secondary <- recode(a13$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
school grads
3044 a13$secondary <- as.factor(a13$secondary)
3045
3046 a13$tertiary <- as.numeric(as.character(a13$A5_15))
3047 a13$tertiary <- recode(a13$tertiary, "0" = "0", "1"="2", "2"="2",
3048                             "3"="2", "4"="2", .default = "3") #0 = no tertiary, 3= tertiary
3049 a13$tertiary <- as.factor(a13$tertiary)
3050
3051 a13 <- mutate(a13, edu = ifelse(secondary == "1" & tertiary == "0", 1,
3052                                ifelse(secondary == "2" & tertiary == "0", 2,
3053                                ifelse(tertiary == "2", 2,

```



```

3054         ifelse(tertiary == "3", 3, 1))))))
3055
3056 a13$edu <- as.factor(a13$edu)
3057 a13$edu <- factor(a13$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
3058
3059 prop_frame(a13, "edu")
3060
3061
3062 ## Deprivation Quintile
3063 names(a13)[names(a13) == "nzdep13_quintile"] <- "dep"
3064
3065 prop_frame(a13, "dep")
3066
3067
3068 ## Fruit and Vegetable
3069 summary(a13$fruit)
3070 summary(a13$veges)
3071 a13$fruit[a13$fruit == "K"] <- NA
3072 a13$fruit[a13$fruit == "R"] <- NA
3073 a13$veges[a13$veges == "K"] <- NA
3074 a13$veges[a13$veges == "R"] <- NA
3075
3076 a13$fruit <- factor(a13$fruit, labels=c("No", "Yes"))
3077 a13$veges <- factor(a13$veges, labels=c("No", "Yes"))
3078
3079 prop_frame(a13, "fruit")
3080 prop_frame(a13, "veges")
3081
3082
3083 ## Smoking Status
3084 a13 <- a13 %>%
3085   mutate(smoke = ifelse(current_smoker == 1 & ex_smoker == 0, 3,
3086     ifelse(current_smoker == 0 & ex_smoker == 1, 2,
3087       ifelse(current_smoker == "K", NA,
3088         ifelse(current_smoker == "R", NA,
3089           ifelse(ex_smoker == "K", NA,
3090             ifelse(ex_smoker == "R", NA, 1)))))))
3091
3092 a13$smoke <- as.factor(a13$smoke)
3093 a13$smoke <- factor(a13$smoke, labels=c("Non Smoker", "Ex Smoker", "Current Smoker"))

```

```

3094
3095 prop_frame(a13, "smoke")
3096
3097
3098 ## Alcohol Problem
3099 a13$haz_drinker_all[a13$haz_drinker_all == "K"] <- NA
3100 a13$haz_drinker_all[a13$haz_drinker_all == "R"] <- NA
3101
3102 a13$haz_drinker_all <- factor(a13$haz_drinker_all, labels=c("No Alcohol Problem",
3103                                                         "Alcohol Problem"))
3104
3105 prop_frame(a13, "haz_drinker_all")
3106
3107
3108 ## Physical Activity
3109 a13$active[a13$active == "K"] <- NA
3110 a13$active[a13$active == "R"] <- NA
3111
3112 a13$active <- factor(a13$active, labels=c("Not Active", "Active"))
3113
3114 prop_frame(a13, "active")
3115
3116
3117 ## Sedentary Lifestyle
3118 a13$sedentary[a13$sedentary == "K"] <- NA
3119 a13$sedentary[a13$sedentary == "R"] <- NA
3120
3121 a13$sedentary <- factor(a13$sedentary, labels=c("Not Sedentary", "Sedentary"))
3122
3123 prop_frame(a13, "sedentary")
3124
3125
3126 # Difficulty Climbing Several Steps of Stairs
3127 summary(a13$A4_03)
3128 a13$A4_03[a13$A4_03 == "."] <- NA
3129 a13$A4_03[a13$A4_03 == ".K"] <- NA
3130 a13$A4_03[a13$A4_03 == ".R"] <- NA
3131
3132 a13$stair <- factor(a13$A4_03, labels=c("A Lot Difficult", "A Little Difficult",
3133                                         "No Difficulty"))

```

```

3134
3135 prop_frame(a13, "stair")
3136
3137
3138
3139 ##### NZHS 2013/14 – child #####
3140
3141 summary(HS13C$bmi) # 981 U, 140 R
3142 HS13C$bmscale <- as.numeric(as.character(HS13C$bmi))
3143
3144 HS13C$gender <- as.factor(HS13C$gender)
3145
3146 # Recode BMI category based on IOTF cut-offs
3147 HS13C <- HS13C %>%
3148   mutate(under =
3149     ifelse(gender == 0 & age == 2 & bmscale <15.24, 0,
3150     ifelse(gender == 0 & age == 3 & bmscale <14.83, 0,
3151     ifelse(gender == 0 & age == 4 & bmscale <14.51, 0,
3152     ifelse(gender == 0 & age == 5 & bmscale <14.26, 0,
3153     ifelse(gender == 0 & age == 6 & bmscale <14.06, 0,
3154     ifelse(gender == 0 & age == 7 & bmscale <14.00, 0,
3155     ifelse(gender == 0 & age == 8 & bmscale <14.13, 0,
3156     ifelse(gender == 0 & age == 9 & bmscale <14.36, 0,
3157     ifelse(gender == 0 & age == 10 & bmscale <14.63, 0,
3158     ifelse(gender == 0 & age == 11 & bmscale <14.96, 0,
3159     ifelse(gender == 0 & age == 12 & bmscale <15.36, 0,
3160     ifelse(gender == 0 & age == 13 & bmscale <15.84, 0,
3161     ifelse(gender == 0 & age == 14 & bmscale <16.39, 0,
3162     ifelse(gender == 0 & age == 15 & bmscale <16.98, 0,
3163     ifelse(gender == 0 & age == 16 & bmscale <17.53, 0,
3164     ifelse(gender == 0 & age == 17 & bmscale <18.04, 0,
3165     ifelse(gender == 1 & age == 2 & bmscale <14.96, 0,
3166     ifelse(gender == 1 & age == 3 & bmscale <14.60, 0,
3167     ifelse(gender == 1 & age == 4 & bmscale <14.30, 0,
3168     ifelse(gender == 1 & age == 5 & bmscale <14.04, 0,
3169     ifelse(gender == 1 & age == 6 & bmscale <13.85, 0,
3170     ifelse(gender == 1 & age == 7 & bmscale <13.83, 0,
3171     ifelse(gender == 1 & age == 8 & bmscale <14.00, 0,
3172     ifelse(gender == 1 & age == 9 & bmscale <14.26, 0,
3173     ifelse(gender == 1 & age == 10 & bmscale <14.58, 0,

```

```

3174 ifelse(gender == 1 & age == 11 & bmiscale <15.03, 0,
3175 ifelse(gender == 1 & age == 12 & bmiscale <15.59, 0,
3176 ifelse(gender == 1 & age == 13 & bmiscale <16.23, 0,
3177 ifelse(gender == 1 & age == 14 & bmiscale <16.86, 0,
3178 ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
3179 ifelse(gender == 1 & age == 16 & bmiscale <17.90, 0,
3180 ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
3181 ifelse(age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))
3182
3183 HS13C <- HS13C %>%
3184 mutate(average =
3185 ifelse(gender == 0 & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
3186 ifelse(gender == 0 & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
3187 ifelse(gender == 0 & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
3188 ifelse(gender == 0 & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
3189 ifelse(gender == 0 & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
3190 ifelse(gender == 0 & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
3191 ifelse(gender == 0 & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
3192 ifelse(gender == 0 & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
3193 ifelse(gender == 0 & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
3194 ifelse(gender == 0 & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
3195 ifelse(gender == 0 & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
3196 ifelse(gender == 0 & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
3197 ifelse(gender == 0 & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
3198 ifelse(gender == 0 & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
3199 ifelse(gender == 0 & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
3200 ifelse(gender == 0 & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
3201 ifelse(gender == 1 & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
3202 ifelse(gender == 1 & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
3203 ifelse(gender == 1 & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
3204 ifelse(gender == 1 & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
3205 ifelse(gender == 1 & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
3206 ifelse(gender == 1 & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
3207 ifelse(gender == 1 & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
3208 ifelse(gender == 1 & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
3209 ifelse(gender == 1 & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
3210 ifelse(gender == 1 & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
3211 ifelse(gender == 1 & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
3212 ifelse(gender == 1 & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
3213 ifelse(gender == 1 & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,

```

```

3214 ifelse(gender == 1 & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
3215 ifelse(gender == 1 & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
3216 ifelse(gender == 1 & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
3217 ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA))))))))))))))))))))))))))))))
3218
3219 HS13C <- HS13C %>%
3220 mutate(over=
3221 ifelse(gender == 0 & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
3222 ifelse(gender == 0 & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
3223 ifelse(gender == 0 & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
3224 ifelse(gender == 0 & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
3225 ifelse(gender == 0 & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
3226 ifelse(gender == 0 & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
3227 ifelse(gender == 0 & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
3228 ifelse(gender == 0 & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
3229 ifelse(gender == 0 & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
3230 ifelse(gender == 0 & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
3231 ifelse(gender == 0 & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
3232 ifelse(gender == 0 & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
3233 ifelse(gender == 0 & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
3234 ifelse(gender == 0 & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
3235 ifelse(gender == 0 & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
3236 ifelse(gender == 0 & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
3237 ifelse(gender == 1 & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
3238 ifelse(gender == 1 & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
3239 ifelse(gender == 1 & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
3240 ifelse(gender == 1 & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
3241 ifelse(gender == 1 & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
3242 ifelse(gender == 1 & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
3243 ifelse(gender == 1 & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
3244 ifelse(gender == 1 & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
3245 ifelse(gender == 1 & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
3246 ifelse(gender == 1 & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
3247 ifelse(gender == 1 & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
3248 ifelse(gender == 1 & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
3249 ifelse(gender == 1 & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
3250 ifelse(gender == 1 & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
3251 ifelse(gender == 1 & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
3252 ifelse(gender == 1 & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
3253 ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))

```

```

3254
3255 HS13C <- HS13C %>%
3256   mutate(obes=
3257     ifelse(gender == 0 & age == 2 & bmiscale >=19.99, 3,
3258     ifelse(gender == 0 & age == 3 & bmiscale >=19.50, 3,
3259     ifelse(gender == 0 & age == 4 & bmiscale >=19.23, 3,
3260     ifelse(gender == 0 & age == 5 & bmiscale >=19.27, 3,
3261     ifelse(gender == 0 & age == 6 & bmiscale >=19.76, 3,
3262     ifelse(gender == 0 & age == 7 & bmiscale >=20.59, 3,
3263     ifelse(gender == 0 & age == 8 & bmiscale >=21.56, 3,
3264     ifelse(gender == 0 & age == 9 & bmiscale >=22.71, 3,
3265     ifelse(gender == 0 & age == 10 & bmiscale >=23.96, 3,
3266     ifelse(gender == 0 & age == 11 & bmiscale >=25.07, 3,
3267     ifelse(gender == 0 & age == 12 & bmiscale >=26.02, 3,
3268     ifelse(gender == 0 & age == 13 & bmiscale >=26.87, 3,
3269     ifelse(gender == 0 & age == 14 & bmiscale >=27.64, 3,
3270     ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
3271     ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
3272     ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
3273     ifelse(gender == 1 & age == 2 & bmiscale >=19.81, 3,
3274     ifelse(gender == 1 & age == 3 & bmiscale >=19.38, 3,
3275     ifelse(gender == 1 & age == 4 & bmiscale >=19.16, 3,
3276     ifelse(gender == 1 & age == 5 & bmiscale >=19.20, 3,
3277     ifelse(gender == 1 & age == 6 & bmiscale >=19.61, 3,
3278     ifelse(gender == 1 & age == 7 & bmiscale >=20.39, 3,
3279     ifelse(gender == 1 & age == 8 & bmiscale >=21.44, 3,
3280     ifelse(gender == 1 & age == 9 & bmiscale >=22.66, 3,
3281     ifelse(gender == 1 & age == 10 & bmiscale >=23.97, 3,
3282     ifelse(gender == 1 & age == 11 & bmiscale >=25.25, 3,
3283     ifelse(gender == 1 & age == 12 & bmiscale >=26.47, 3,
3284     ifelse(gender == 1 & age == 13 & bmiscale >=27.57, 3,
3285     ifelse(gender == 1 & age == 14 & bmiscale >=28.42, 3,
3286     ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
3287     ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
3288     ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
3289     ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))
3290
3291
3292 # 0: underweight, 1: average, 2: overweight, 3: obese
3293 HS13C <- HS13C %>%

```

```

3294   mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
3295                     ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
3296                     ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
3297                     ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA))))))
3298
3299 HS13C$bmic <- as.factor(HS13C$bmic)
3300
3301 # Exclude underweight
3302 c13 <- subset(HS13C, !bmic == 0 | is.na(bmic))
3303
3304
3305 ## BMI
3306 c13$bmic <- factor(c13$bmic, labels=c("Average", "Overweight", "Obese"))
3307
3308 prop_frame(c13, "bmic")
3309
3310
3311 ## Sex
3312 c13$gender <- factor(c13$gender, labels=c("Male", "Female"))
3313
3314 prop_frame(c13, "gender")
3315
3316
3317 ## Age
3318 summary(c13$age)
3319
3320
3321 ## Ethnicity
3322 # recode anyone ~ european
3323 c13 = c13 %>%
3324   mutate(euro = ifelse(C4_03_01 == 1, 1, 0))
3325
3326 # recode anyone ~ maori
3327 c13 = c13 %>%
3328   mutate(maori = ifelse(C4_03_02 == 1, 1, 0))
3329
3330 # recode anyone ~ pacific
3331 c13 = c13 %>%
3332   mutate(pacific = ifelse(C4_03_03 == 1, 1,
3333                           ifelse(C4_03_04 == 1, 1,

```

```

3334         ifelse(C4_03_05 == 1, 1,
3335         ifelse(C4_03_06 == 1, 1, 0))))))
3336
3337 # recode anyone ~ asian
3338 c13 = c13 %>%
3339   mutate(asian = ifelse(C4_03_07 == 1, 1,
3340     ifelse(C4_03_08 == 1, 1, 0)))
3341
3342 # recode anyone ~ other
3343 c13 = c13 %>%
3344   mutate(other = ifelse(C4_03_77 == 1, 1, 0))
3345
3346
3347 c13 = c13 %>%
3348   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
3349     Only",
3350     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
3351       Pacific Only",
3352     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
3353       Only",
3354     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
3355       European Only",
3356     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
3357       ,
3358     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
3359       Ethnicities (M)",
3360     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
3361       Ethnicities (M)",
3362     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
3363       ,
3364     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
3365       ,
3366     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
3367       ,
3368     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
3369       Ethnicities (M)",
3370     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
3371       Ethnicities (M)",
3372     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
3373       ,

```



```

3361         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
3362         ,
3363         ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
3364         ,
3365         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
          Ethnicities (M)",
3366         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
          Ethnicities (M)",
3367         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
          Ethnicities (M)",
3368         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
3369         ,
3370         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
3371         ,
3372         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
3373         ,
3374         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
          Ethnicities (M)",
3375         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
          Ethnicities (M)",
3376         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
          Ethnicities (M)",
3377         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
3378         ,
3379         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
          Ethnicities (M)",
3380         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
          Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))))))
3381
3382 c13$eth_count <- as.factor(c13$eth_count)
3383
3384 prop_frame(c13, "eth_count")

```

```

3383
3384
3385 ## Household Income
3386 # combined adult to child data set to match income and education data
3387 ac13 <- left_join(select(c13, HHID, C4_17, C4_20, C4_21, CQ1, CQ2),
3388                   select(a13, HHID, A5_24, A5_14, A5_15), by="HHID")
3389
3390 summary(ac13$A5_24)
3391 ac13$hhinc1 <- ordered(ac13$A5_24, levels=c("4","5","6","7","8","9","10","11","12",
3392                                             "13","14","15","16"))
3393 summary(ac13$C4_17)
3394 ac13$hhinc2 <- ordered(ac13$C4_17, levels=c("4","5","6","7","8","9","10","11","12",
3395                                             "13","14","15","16"))
3396
3397 ac13$hhinc1 <- as.numeric(as.character(ac13$hhinc1))
3398 ac13$hhinc2 <- as.numeric(as.character(ac13$hhinc2))
3399
3400 ac13$hhinc1 <- cut(ac13$hhinc1, breaks=c(0,5,10,13,16))
3401 ac13$hhinc2 <- cut(ac13$hhinc2, breaks=c(0,5,10,13,16))
3402
3403 ac13$hhinc1 <- factor(ac13$hhinc1, labels=c(1,2,3,4))
3404 ac13$hhinc2 <- factor(ac13$hhinc2, labels=c(1,2,3,4))
3405
3406 ac13 <- mutate(ac13, hhinc= ifelse(hhinc1 == 1 & CQ1 == 1, 1,
3407                                   ifelse(hhinc1 == 2 & CQ1 == 1, 2,
3408                                           ifelse(hhinc1 == 3 & CQ1 == 1, 3,
3409                                                   ifelse(hhinc1 == 4 & CQ1 == 1, 4,
3410                                                           ifelse(hhinc2 == 1 & CQ1 == 2, 1,
3411                                                                 ifelse(hhinc2 == 2 & CQ1 == 2, 2,
3412                                                                       ifelse(hhinc2 == 3 & CQ1 == 2, 3,
3413                                                                           ifelse(hhinc2 == 4 & CQ1 == 2, 4, NA))))))))))
3414
3415 ac13$hhinc <- as.factor(ac13$hhinc)
3416
3417 # recode back into child data set
3418 ac13$hhinc <- factor(ac13$hhinc, labels=c("<=$15,000",
3419                                           "$15,001-$40,000", "$40,001-$70,000", ">$70,000"))
3420 c13$hhinc <- ac13$hhinc
3421
3422 prop_frame(c13, "hhinc")

```

```

3423
3424
3425 ## Education
3426 ac13$cedu1 <- as.numeric(as.character(ac13$C4_20))
3427 ac13$aedu1 <- as.numeric(as.character(ac13$A5_14))
3428
3429 ac13 <- mutate(ac13, edu1 = ifelse(is.na(cedu1), aedu1,
3430                                   ifelse(is.na(aedu1), cedu1,
3431                                           ifelse(cedu1 >= aedu1, cedu1,
3432                                                   ifelse(aedu1 >= cedu1, aedu1, NA))))))
3433
3434 ac13$edu1 <- as.factor(ac13$edu1)
3435 summary(ac13$edu1)
3436
3437 # coding discrepancies, decided to use the highest qualification in the household
3438 ac13$secondary <- as.numeric(as.character(ac13$edu1))
3439 ac13$secondary <- recode(ac13$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
3440                                school grads
3441 ac13$secondary <- as.factor(ac13$secondary)
3442
3442 summary(ac13$C4_21) # used different coding than the adult set
3443 summary(ac13$A5_15)
3444
3445 ac13$C4_21[ac13$C4_21 == "77"] <- NA
3446 ac13$A5_15[ac13$A5_15 == "77"] <- NA
3447
3448 ac13$cedu2 <- as.numeric(as.character(ac13$C4_21))-1 # to match the coding in the adult set
3449 ac13$aedu2 <- as.numeric(as.character(ac13$A5_15))
3450
3451 ac13 <- mutate(ac13, edu2 = ifelse(is.na(cedu2), aedu2,
3452                                   ifelse(is.na(aedu2), cedu2,
3453                                           ifelse(aedu2 >= cedu2, aedu2,
3454                                                   ifelse(cedu2 >= aedu2, cedu2, NA))))))
3455
3456 ac13$edu2 <- as.factor(ac13$edu2)
3457 summary(ac13$edu2)
3458
3459
3460 ac13$tertiary <- as.numeric(as.character(ac13$edu2))
3461 ac13$tertiary <- recode(ac13$tertiary, "0" = "0", "1" = "2", "2" = "2",

```

```

3462         "3"="2", "4"="2", .default = "3") #0 = no tertiary , 3= tertiary
3463 ac13$tertiary <- as.factor(ac13$tertiary)
3464
3465 # recode into one variable , prioritise highest education
3466 ac13 <- mutate(ac13, edu = ifelse(secondary == "1" & tertiary == "0", 1,
3467                                   ifelse(secondary == "2" & tertiary == "0", 2,
3468                                           ifelse(tertiary == "2", 2,
3469                                                   ifelse(tertiary == "3", 3, 1))))))
3470
3471 ac13$edu <- factor(ac13$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
3472
3473 # move them back into child data set
3474 c13$edu <- ac13$edu
3475
3476 prop_frame(c13, "edu")
3477
3478 # remove ac13 from memory
3479 rm(ac13)
3480
3481
3482 ## Deprivation Quintile
3483 names(c13)[names(c13) == "nzdep13_quintile"] <- "dep"
3484
3485 prop_frame(c13, "dep")
3486
3487
3488 ## Fruit and Vegetable
3489 c13$fruit <- c13$kidfruit_2serves
3490 c13$veges <- c13$kidveges_new
3491
3492 c13$fruit[c13$fruit == "K"] <- NA
3493 c13$fruit[c13$fruit == "X"] <- NA
3494 c13$veges[c13$veges == "K"] <- NA
3495 c13$veges[c13$veges == "R"] <- NA
3496 c13$veges[c13$veges == "X"] <- NA
3497
3498 c13$fruit <- factor(c13$fruit, labels=c("No", "Yes"))
3499 c13$veges <- factor(c13$veges, labels=c("No", "Yes"))
3500
3501 prop_frame(c13, "fruit")

```

```

3502 prop_frame(c13, "veges")
3503
3504
3505 ### Soft Drink and Fast Food
3506 summary(c13$C3_09) # 2 NA, 16 K
3507 c13$softd <- as.numeric(as.character(c13$C3_09)) # force non-numeric into NA
3508 c13$softd <- cut(c13$softd, c(0,1,2,4,Inf), labels=c("0/week",
3509             "1/week", "2-3/week", "4+/week"), right=FALSE)
3510
3511 summary(c13$C3_10)
3512 c13$fastf <- as.numeric(as.character(c13$C3_10)) # force non-numeric into NA
3513 c13$fastf <- cut(c13$fastf, c(0,1,2,4,Inf), labels=c("0/week",
3514             "1/week", "2-3/week", "4+/week"), right=FALSE)
3515
3516 prop_frame(c13, "softd")
3517 prop_frame(c13, "fastf")
3518
3519
3520
3521
3522 ##### NZHS 2014/15 – adult #####
3523
3524 summary(HS14A$bmi) # 527 R, 201 P
3525
3526 HS14A$gender <- as.factor(HS14A$gender)
3527
3528 HS14A$bmiscale <- as.numeric(as.character(HS14A$bmi))
3529
3530 # recode BMI category using IOTF cut-offs
3531 HS14A <- HS14A %>%
3532   mutate(bmic = ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
3533     ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
3534     ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
3535     ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
3536     ifelse(gender == 1 & age == 16 & bmiscale <17.9, 0,
3537     ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
3538     ifelse(age >= 18 & bmiscale <18.5, 0,
3539     ifelse(gender == 0 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
3540     ifelse(gender == 0 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
3541     ifelse(gender == 0 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,

```

```

3542         ifelse(gender == 1 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
3543         ifelse(gender == 1 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
3544         ifelse(gender == 1 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
3545         ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
3546         ifelse(gender == 0 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
3547         ifelse(gender == 0 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
3548         ifelse(gender == 0 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
3549         ifelse(gender == 1 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
3550         ifelse(gender == 1 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
3551         ifelse(gender == 1 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
3552         ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
3553         ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
3554         ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
3555         ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
3556         ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
3557         ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
3558         ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
3559         ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))))))
3560
3561 HS14A$bmic <- as.factor(HS14A$bmic)
3562 summary(HS14A$bmic)
3563
3564 # exclude underweight
3565 a14 <- subset(HS14A, !bmic == 0 | is.na(bmic))
3566
3567
3568 ## Migration Status
3569 summary(a14$A5_06)
3570
3571 a14$native <- as.numeric(as.character(a14$A5_06))
3572 a14$native <- cut(a14$native, c(0,2005,Inf), right=FALSE)
3573 summary(a14$native)
3574
3575 a14$native <- factor(a14$native, labels= c("Native","Migrant"))
3576 a14$native[is.na(a14$native)] <- "Native" # convert all NA into native
3577
3578 prop_frame(a14, "native")
3579
3580
3581 ## BMI

```

```

3582 a14$bmic <- factor(a14$bmic, labels=c("Average", "Overweight", "Obese"))
3583
3584 prop_frame(a14, "bmic")
3585
3586
3587 ## Sex
3588 a14$gender <- factor(a14$gender, labels=c("Male", "Female"))
3589
3590 prop_frame(a14, "gender")
3591
3592
3593 ## Age
3594 summary(a14$age)
3595
3596
3597 ## Ethnicity
3598 # recode anyone ~ european
3599 a14 = a14 %>%
3600   mutate(euro = ifelse(A5_03_01 == 1, 1, 0))
3601
3602 # recode anyone ~ maori
3603 a14 = a14 %>%
3604   mutate(maori = ifelse(A5_03_02 == 1, 1, 0))
3605
3606 # recode anyone ~ pacific
3607 a14 = a14 %>%
3608   mutate(pacific = ifelse(A5_03_03 == 1, 1,
3609     ifelse(A5_03_04 == 1, 1,
3610       ifelse(A5_03_05 == 1, 1,
3611         ifelse(A5_03_06 == 1, 1, 0))))))
3612
3613 # recode anyone ~ asian (Indian and Chinese)
3614 a14 = a14 %>%
3615   mutate(asian = ifelse(A5_03_07 == 1, 1,
3616     ifelse(A5_03_08 == 1, 1, 0)))
3617
3618 # recode anyone ~ other
3619 a14 = a14 %>%
3620   mutate(other = ifelse(A5_03_77 == 1, 1, 0))
3621

```

```

3622 a14 = a14 %%%
3623 mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
    Only",
3624
    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
        Pacific Only",
3625
    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
        Only",
3626
    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
        European Only",
3627
    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
        ,
3628
    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
        Ethnicities (M)",
3629
    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
        Ethnicities (M)",
3630
    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
        ,
3631
    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
        ,
3632
    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
        ,
3633
    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
        Ethnicities (M)",
3634
    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
        Ethnicities (M)",
3635
    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
        ,
3636
    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
        ,
3637
    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
        ,
3638
    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
        Ethnicities (M)",
3639
    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
        Ethnicities (M)",
3640
    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
        Ethnicities (M)",
3641
    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
        ,

```



```

3642         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
3643         ,
3644         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
3645         ,
3646         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
3647             Ethnicities (M)",
3648         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
3649             Ethnicities (M)",
3650         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
3651             Ethnicities (M)",
3652         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
3653         ,
3654         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
3655             Ethnicities (M)",
3656         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
3657             Ethnicities (M)",
3658         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
3659             Ethnicities (M)",
3660         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
3661         ,
3662         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
3663             Ethnicities (M)",
3664         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
3665             Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))))))
3666
3667 a14$eth_count <- as.factor(a14$eth_count)
3668
3669 prop_frame(a14, "eth_count")
3670
3671
3672 ## Household Income
3673 summary(a14$A5_24)
3674 a14$hhinc <- ordered(a14$A5_24, levels=c("4", "5", "6", "7", "8", "9", "10", "11", "12",
3675     "13", "14", "15", "16"))
3676
3677 a14$hhinc <- as.numeric(as.character(a14$hhinc)) # force non-numeric into NA
3678 a14$hhinc <- cut(a14$hhinc, breaks=c(0,5,10,13,16))
3679 a14$hhinc <- factor(a14$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000", ">$70,000"))
3680
3681 prop_frame(a14, "hhinc")

```

```

3670
3671
3672 ## Education
3673 a14$secondary <- as.numeric(as.character(a14$A5_14)) # force non-numeric into NA
3674 a14$secondary <- recode(a14$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
3675 a14$secondary <- as.factor(a14$secondary)
3676
3677 a14$tertiary <- as.numeric(as.character(a14$A5_15)) # force non-numeric into NA
3678 a14$tertiary <- recode(a14$tertiary, "0" = "0", "1"="2", "2"="2",
3679                        "3"="2", "4"="2", .default = "3") #0 = no tertiary, 3= tertiary
3680 a14$tertiary <- as.factor(a14$tertiary)
3681
3682 a14 <- mutate(a14, edu = ifelse(secondary == "1" & tertiary == "0", 1,
3683                                ifelse(secondary == "2" & tertiary == "0", 2,
3684                                ifelse(tertiary == "2", 2,
3685                                ifelse(tertiary == "3", 3, 1))))))
3686
3687 a14$edu <- as.factor(a14$edu)
3688
3689 a14$edu <- factor(a14$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
3690
3691 prop_frame(a14, "edu")
3692
3693
3694 ## Deprivation Quintile
3695 names(a14)[names(a14) == "nzdep13_quintile"] <- "dep"
3696
3697 prop_frame(a14, "dep")
3698
3699
3700 ## Fruit and Vegetable
3701 a14$fruit[a14$fruit == "K"] <- NA
3702 a14$veges[a14$veges == "K"] <- NA
3703 a14$fruit <- factor(a14$fruit, labels=c("No", "Yes"))
3704 a14$veges <- factor(a14$veges, labels=c("No", "Yes"))
3705
3706 prop_frame(a14, "fruit")
3707 prop_frame(a14, "veges")
3708

```

```

3709
3710 ## Alcohol Problem
3711 a14$haz_drinker_all[a14$haz_drinker_all == "K"] <- NA
3712 a14$haz_drinker_all[a14$haz_drinker_all == "R"] <- NA
3713
3714 a14$haz_drinker_all <- factor(a14$haz_drinker_all, labels=c("No Alcohol Problem",
3715                                                             "Alcohol Problem"))
3716
3717 prop_frame(a14, "haz_drinker_all")
3718
3719
3720 ## Smoking Status
3721 a14 <- a14 %>%
3722   mutate(smoke = ifelse(current_smoker == 1 & ex_smoker == 0, 3,
3723                         ifelse(current_smoker == 0 & ex_smoker == 1, 2,
3724                               ifelse(current_smoker == "K", NA,
3725                                     ifelse(current_smoker == "R", NA,
3726                                             ifelse(ex_smoker == "K", NA,
3727                                                   ifelse(ex_smoker == "R", NA, 1)))))))
3728
3729 a14$smoke <- as.factor(a14$smoke)
3730 a14$smoke <- factor(a14$smoke, labels=c("Non Smoker", "Ex Smoker", "Current Smoker"))
3731 summary(a14$smoke)
3732
3733 prop_frame(a14, "smoke")
3734
3735
3736 ## Physical ACtivity
3737 a14$active[a14$active == "K"] <- NA
3738 a14$active[a14$active == "R"] <- NA
3739 a14$active <- factor(a14$active, labels=c("Not Active", "Active"))
3740
3741 prop_frame(a14, "active")
3742
3743
3744 ## Sedentary Lifestyle
3745 a14$sedentary[a14$sedentary == "K"] <- NA
3746 a14$sedentary[a14$sedentary == "R"] <- NA
3747 a14$sedentary <- factor(a14$sedentary, labels=c("Not Sedentary", "Sedentary"))
3748

```

```

3749 prop_frame(a14, "sedentary")
3750
3751
3752 ## Urban/Rural Area
3753 a14$UA <- recode(a14$UrbanArea2015Class, "M"="Urban", "S"="Urban", .default="Rural")
3754
3755 prop_frame(a14, "UA")
3756
3757
3758 ## Difficulty Climbing Several Steps of Stairs
3759 summary(a14$A4_03)
3760 a14$A4_03[a14$A4_03 == ".K"] <- NA
3761 a14$A4_03[a14$A4_03 == ".R"] <- NA
3762
3763 a14$stair <- factor(a14$A4_03, labels=c("A Lot Difficult", "A Little Difficult",
3764                                         "No Difficulty"))
3765
3766 prop_frame(a14, "stair")
3767
3768
3769
3770 ##### NZHS 2014/15 – child #####
3771
3772 summary(HS14C$bmi) # 981 U, 276 R, 6 O
3773
3774 HS14C$bmscale <- as.numeric(as.character(HS14C$bmi))
3775
3776 HS14C$gender <- as.factor(HS14C$gender)
3777
3778 # recode into BMI Category based on IOTF cut-offs
3779 HS14C <- HS14C %>%
3780   mutate(under =
3781     ifelse(gender == 0 & age == 2 & bmscale <15.24, 0,
3782     ifelse(gender == 0 & age == 3 & bmscale <14.83, 0,
3783     ifelse(gender == 0 & age == 4 & bmscale <14.51, 0,
3784     ifelse(gender == 0 & age == 5 & bmscale <14.26, 0,
3785     ifelse(gender == 0 & age == 6 & bmscale <14.06, 0,
3786     ifelse(gender == 0 & age == 7 & bmscale <14.00, 0,
3787     ifelse(gender == 0 & age == 8 & bmscale <14.13, 0,
3788     ifelse(gender == 0 & age == 9 & bmscale <14.36, 0,

```

```

3789   ifelse(gender == 0 & age == 10 & bmiscale <14.63, 0,
3790   ifelse(gender == 0 & age == 11 & bmiscale <14.96, 0,
3791   ifelse(gender == 0 & age == 12 & bmiscale <15.36, 0,
3792   ifelse(gender == 0 & age == 13 & bmiscale <15.84, 0,
3793   ifelse(gender == 0 & age == 14 & bmiscale <16.39, 0,
3794   ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
3795   ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
3796   ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
3797   ifelse(gender == 1 & age == 2 & bmiscale <14.96, 0,
3798   ifelse(gender == 1 & age == 3 & bmiscale <14.60, 0,
3799   ifelse(gender == 1 & age == 4 & bmiscale <14.30, 0,
3800   ifelse(gender == 1 & age == 5 & bmiscale <14.04, 0,
3801   ifelse(gender == 1 & age == 6 & bmiscale <13.85, 0,
3802   ifelse(gender == 1 & age == 7 & bmiscale <13.83, 0,
3803   ifelse(gender == 1 & age == 8 & bmiscale <14.00, 0,
3804   ifelse(gender == 1 & age == 9 & bmiscale <14.26, 0,
3805   ifelse(gender == 1 & age == 10 & bmiscale <14.58, 0,
3806   ifelse(gender == 1 & age == 11 & bmiscale <15.03, 0,
3807   ifelse(gender == 1 & age == 12 & bmiscale <15.59, 0,
3808   ifelse(gender == 1 & age == 13 & bmiscale <16.23, 0,
3809   ifelse(gender == 1 & age == 14 & bmiscale <16.86, 0,
3810   ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
3811   ifelse(gender == 1 & age == 16 & bmiscale <17.90, 0,
3812   ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
3813   ifelse(age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))
3814
3815 HS14C <- HS14C %>%
3816   mutate(average =
3817     ifelse(gender == 0 & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
3818     ifelse(gender == 0 & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
3819     ifelse(gender == 0 & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
3820     ifelse(gender == 0 & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
3821     ifelse(gender == 0 & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
3822     ifelse(gender == 0 & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
3823     ifelse(gender == 0 & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
3824     ifelse(gender == 0 & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
3825     ifelse(gender == 0 & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
3826     ifelse(gender == 0 & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
3827     ifelse(gender == 0 & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
3828     ifelse(gender == 0 & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,

```

```

3829 ifelse(gender == 0 & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
3830 ifelse(gender == 0 & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
3831 ifelse(gender == 0 & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
3832 ifelse(gender == 0 & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
3833 ifelse(gender == 1 & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
3834 ifelse(gender == 1 & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
3835 ifelse(gender == 1 & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
3836 ifelse(gender == 1 & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
3837 ifelse(gender == 1 & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
3838 ifelse(gender == 1 & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
3839 ifelse(gender == 1 & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
3840 ifelse(gender == 1 & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
3841 ifelse(gender == 1 & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
3842 ifelse(gender == 1 & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
3843 ifelse(gender == 1 & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
3844 ifelse(gender == 1 & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
3845 ifelse(gender == 1 & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
3846 ifelse(gender == 1 & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
3847 ifelse(gender == 1 & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
3848 ifelse(gender == 1 & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
3849 ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA))))))))))))))))))))))))))))))
3850
3851 HS14C <- HS14C %>%
3852 mutate(over=
3853 ifelse(gender == 0 & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
3854 ifelse(gender == 0 & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
3855 ifelse(gender == 0 & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
3856 ifelse(gender == 0 & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
3857 ifelse(gender == 0 & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
3858 ifelse(gender == 0 & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
3859 ifelse(gender == 0 & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
3860 ifelse(gender == 0 & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
3861 ifelse(gender == 0 & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
3862 ifelse(gender == 0 & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
3863 ifelse(gender == 0 & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
3864 ifelse(gender == 0 & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
3865 ifelse(gender == 0 & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
3866 ifelse(gender == 0 & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
3867 ifelse(gender == 0 & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
3868 ifelse(gender == 0 & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,

```

```

3869 ifelse(gender == 1 & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
3870 ifelse(gender == 1 & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
3871 ifelse(gender == 1 & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
3872 ifelse(gender == 1 & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
3873 ifelse(gender == 1 & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
3874 ifelse(gender == 1 & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
3875 ifelse(gender == 1 & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
3876 ifelse(gender == 1 & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
3877 ifelse(gender == 1 & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
3878 ifelse(gender == 1 & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
3879 ifelse(gender == 1 & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
3880 ifelse(gender == 1 & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
3881 ifelse(gender == 1 & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
3882 ifelse(gender == 1 & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
3883 ifelse(gender == 1 & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
3884 ifelse(gender == 1 & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
3885 ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))
3886
3887 HS14C <- HS14C %>%
3888 mutate(obes=
3889 ifelse(gender == 0 & age == 2 & bmiscale >=19.99, 3,
3890 ifelse(gender == 0 & age == 3 & bmiscale >=19.50, 3,
3891 ifelse(gender == 0 & age == 4 & bmiscale >=19.23, 3,
3892 ifelse(gender == 0 & age == 5 & bmiscale >=19.27, 3,
3893 ifelse(gender == 0 & age == 6 & bmiscale >=19.76, 3,
3894 ifelse(gender == 0 & age == 7 & bmiscale >=20.59, 3,
3895 ifelse(gender == 0 & age == 8 & bmiscale >=21.56, 3,
3896 ifelse(gender == 0 & age == 9 & bmiscale >=22.71, 3,
3897 ifelse(gender == 0 & age == 10 & bmiscale >=23.96, 3,
3898 ifelse(gender == 0 & age == 11 & bmiscale >=25.07, 3,
3899 ifelse(gender == 0 & age == 12 & bmiscale >=26.02, 3,
3900 ifelse(gender == 0 & age == 13 & bmiscale >=26.87, 3,
3901 ifelse(gender == 0 & age == 14 & bmiscale >=27.64, 3,
3902 ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
3903 ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
3904 ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
3905 ifelse(gender == 1 & age == 2 & bmiscale >=19.81, 3,
3906 ifelse(gender == 1 & age == 3 & bmiscale >=19.38, 3,
3907 ifelse(gender == 1 & age == 4 & bmiscale >=19.16, 3,
3908 ifelse(gender == 1 & age == 5 & bmiscale >=19.20, 3,

```

```

3909   ifelse(gender == 1 & age == 6 & bmiscale >=19.61, 3,
3910   ifelse(gender == 1 & age == 7 & bmiscale >=20.39, 3,
3911   ifelse(gender == 1 & age == 8 & bmiscale >=21.44, 3,
3912   ifelse(gender == 1 & age == 9 & bmiscale >=22.66, 3,
3913   ifelse(gender == 1 & age == 10 & bmiscale >=23.97, 3,
3914   ifelse(gender == 1 & age == 11 & bmiscale >=25.25, 3,
3915   ifelse(gender == 1 & age == 12 & bmiscale >=26.47, 3,
3916   ifelse(gender == 1 & age == 13 & bmiscale >=27.57, 3,
3917   ifelse(gender == 1 & age == 14 & bmiscale >=28.42, 3,
3918   ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
3919   ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
3920   ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
3921   ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))
3922
3923
3924 # 0: underweight, 1: average, 2: overweight, 3: obese
3925 HS14C <- HS14C %>%
3926   mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
3927     ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
3928     ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
3929     ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA))))
3930
3931 HS14C$bmic <- as.factor(HS14C$bmic)
3932 summary(HS14C$bmic)
3933
3934 # exclude underweight
3935 c14<- subset(HS14C, !bmic == 0 | is.na(bmic))
3936
3937
3938 ## BMI
3939 summary(c14$bmiscale)
3940
3941 c14$bmic <- factor(c14$bmic, labels=c("Average", "Overweight", "Obese"))
3942
3943 prop_frame(c14, "bmic")
3944
3945
3946 ## Sex
3947 c14$gender <- factor(c14$gender, labels=c("Male", "Female"))
3948

```



```

3949 prop_frame(c14, "gender")
3950
3951
3952 ## Age
3953 summary(c14$age)
3954
3955
3956 ## Ethnicity
3957 # recode anyone ~ european
3958 c14 = c14 %>%
3959   mutate(euro = ifelse(C4_03_01 == 1, 1, 0))
3960
3961 # recode anyone ~ maori
3962 c14 = c14 %>%
3963   mutate(maori = ifelse(C4_03_02 == 1, 1, 0))
3964
3965 # recode anyone ~ pacific
3966 c14 = c14 %>%
3967   mutate(pacific = ifelse(C4_03_03 == 1, 1,
3968                           ifelse(C4_03_04 == 1, 1,
3969                                   ifelse(C4_03_05 == 1, 1,
3970                                           ifelse(C4_03_06 == 1, 1, 0))))))
3971
3972 # recode anyone ~ asian (Indian and Chinese)
3973 c14 = c14 %>%
3974   mutate(asian = ifelse(C4_03_07 == 1, 1,
3975                         ifelse(C4_03_08 == 1, 1, 0)))
3976
3977 # recode anyone ~ other
3978 c14 = c14 %>%
3979   mutate(other = ifelse(C4_03_77 == 1, 1, 0))
3980
3981
3982 c14 = c14 %>%
3983   mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
3984                               Only",
3985                               ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
3986                                       Pacific Only",
3987                                       ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
3988                                               Only",

```

```

3986         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
3987             European Only",
3988         ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
3989             ,
3990         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
3991             Ethnicities (M)",
3992         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
3993             Ethnicities (M)",
3994         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other"
3995             ,
3996         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other"
3997             ,
3998         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
3999             ,
4000         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
4001             Ethnicities (M)",
4002         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
4003             Ethnicities (M)",
4004         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
4005             ,
4006         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
4007             ,
4008         ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
4009             ,
4010         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
4011             Ethnicities (M)",
4012         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
4013             Ethnicities (M)",
4014         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
4015             Ethnicities (M)",
4016         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
4017             ,
4018         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
4019             ,
4020         ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
4021             ,
4022         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
4023             Ethnicities (M)",
4024         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
4025             Ethnicities (M)",

```

```

4006         ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
           Ethnicities (M)",
4007         ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
           ,
4008         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
           Ethnicities (M)",
4009         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
           Ethnicities (M)",
4010         ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
           Ethnicities (M)",
4011         ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
           ,
4012         ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
           Ethnicities (M)",
4013         ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
           Ethnicities (M)", NA))))))))))))))))))))))))))))))))))))))))))
4014
4015 c14$eth_count <- as.factor(c14$eth_count)
4016
4017 prop_frame(c14, "eth_count")
4018
4019
4020 ## Household Income
4021 # combine adult to child data to match income and education data
4022 ac14 <- left_join(select(c14, HHID, C4_17, C4_20, C4_21),
4023                   select(a14, HHID, A5_24, A5_14, A5_15), by="HHID")
4024
4025 summary(ac14$A5_24)
4026 ac14$hhinc1 <- ordered(ac14$A5_24, levels=c("4","5","6","7","8","9","10","11","12",
4027                                             "13","14","15","16"))
4028 summary(ac14$C4_17)
4029 ac14$hhinc2 <- ordered(ac14$C4_17, levels=c("4","5","6","7","8","9","10","11","12",
4030                                             "13","14","15","16"))
4031
4032 ac14$hhinc1 <- as.numeric(as.character(ac14$hhinc1))
4033 ac14$hhinc2 <- as.numeric(as.character(ac14$hhinc2))
4034
4035 ac14 <- mutate(ac14, inc= ifelse(is.na(hhinc2), hhinc1,
4036                                 ifelse(is.na(hhinc1), hhinc2, hhinc1)))
4037

```

```

4038 ac14$hhinc <- cut(ac14$inc, breaks=c(0,5,10,13,16))
4039 ac14$hhinc <- factor(ac14$hhinc, labels=c("<=$15,000", "$15,001-$40,000", "$40,001-$70,000", ">$70,000"
    ))
4040
4041 ac14$hhinc <- as.factor(ac14$hhinc)
4042 summary(ac14$hhinc)
4043
4044 # recode back into original data set
4045 c14$hhinc <- ac14$hhinc
4046
4047 prop_frame(c14, "hhinc")
4048
4049
4050 ## Education
4051 ac14$cedu1 <- as.numeric(as.character(ac14$C4_20))
4052 ac14$aedu1 <- as.numeric(as.character(ac14$A5_14))
4053
4054 ac14 <- mutate(ac14, edu1 = ifelse(is.na(cedu1), aedu1,
4055                                   ifelse(is.na(aedu1), cedu1,
4056                                           ifelse(cedu1 >= aedu1, cedu1,
4057                                                   ifelse(aedu1 >= cedu1, aedu1, NA))))))
4058
4059 ac14$edu1 <- as.factor(ac14$edu1)
4060
4061 ac14$secondary <- as.numeric(as.character(ac14$edu1))
4062 ac14$secondary <- recode(ac14$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
4063 ac14$secondary <- as.factor(ac14$secondary)
4064 summary(ac14$secondary)
4065
4066
4067 ac14$C4_21[ac14$C4_21 == "77"] <- NA
4068 ac14$A5_15[ac14$A5_15 == "77"] <- NA
4069
4070 ac14$cedu2 <- as.numeric(as.character(ac14$C4_21))
4071 ac14$aedu2 <- as.numeric(as.character(ac14$A5_15))
4072
4073 ac14 <- mutate(ac14, edu2 = ifelse(is.na(cedu2), aedu2,
4074                                   ifelse(is.na(aedu2), cedu2,
4075                                           ifelse(cedu2 >= aedu2, cedu2,

```

```

4076         ifelse(aedu2 >= cedu2, aedu2, NA))))))
4077
4078 ac14$edu2 <- as.factor(ac14$edu2)
4079
4080 ac14$tertiary <- as.numeric(as.character(ac14$edu2))
4081 ac14$tertiary <- recode(ac14$tertiary, "0" = "0", "1"="2", "2"="2",
4082         "3"="2", "4"="2", .default = "3") #0 = no tertiary, 3= tertiary
4083 ac14$tertiary <- as.factor(ac14$tertiary)
4084
4085 # recode into one variable, prioritise highest education
4086 ac14 <- mutate(ac14, edu = ifelse(secondary == "1" & tertiary == "0", 1,
4087         ifelse(secondary == "2" & tertiary == "0", 2,
4088         ifelse(tertiary == "2", 2,
4089         ifelse(tertiary == "3", 3, 1))))))
4090 summary(ac14$edu)
4091
4092 # recode back into child data set
4093 c14$edu <- factor(ac14$edu, labels=c("No Qualification", "Secondary", "Tertiary"))
4094
4095 prop_frame(c14, "edu")
4096
4097 # remove ac14 from memory
4098 rm(ac14)
4099
4100
4101 ## Deprivation Quintile
4102 names(c14)[names(c14) == "nzdep13_quintile"] <- "dep"
4103
4104 prop_frame(c14, "dep")
4105
4106
4107 ## Fruit and Vegetable
4108 c14$kidfruit_2serves[c14$kidfruit_2serves == "K"] <- NA
4109 c14$kidfruit_2serves[c14$kidfruit_2serves == "X"] <- NA
4110 c14$kidveges_new[c14$kidveges_new == "K"] <- NA
4111 c14$kidveges_new[c14$kidveges_new == "R"] <- NA
4112 c14$kidveges_new[c14$kidveges_new == "X"] <- NA
4113
4114 c14$fruit <- factor(c14$kidfruit_2serves, labels=c("No", "Yes"))
4115 c14$veges <- factor(c14$kidveges_new, labels=c("No", "Yes"))

```

```

4116
4117 prop_frame(c14, "fruit")
4118 prop_frame(c14, "veges")
4119
4120
4121 ## Urban/Rural Area
4122 c14$UA <- recode(c14$UrbanArea2015Class, "M"="Urban", "S"="Urban", .default="Rural")
4123
4124 prop_frame(c14, "UA")
4125
4126
4127 ## Soft Drink and Fast Food
4128 summary(c14$C3_09) # 21 K
4129 c14$softd <- as.numeric(as.character(c14$C3_09))
4130 c14$softd <- cut(c14$softd, c(0,1,2,4,Inf), labels=c("0/week", "1/week",
4131             "2-3/week", "4+/week"), right=FALSE)
4132
4133 summary(c14$C3_10) # 11 K
4134 c14$fastf <- as.numeric(as.character(c14$C3_10))
4135 c14$fastf <- cut(c14$fastf, c(0,1,2,4,Inf), labels=c("0/week", "1/week",
4136             "2-3/week", "4+/week"), right=FALSE)
4137
4138 prop_frame(c14, "softd")
4139 prop_frame(c14, "fastf")

```

Listing A.2: The NZHSs Data Analyses

```
1
2 ##### Data Analyses – New Zealand Health Surveys
3
4 ##### Merge the NZHSs together #####
5 # Merge adult and child data, then split them into two groups: 2–17 years old and 18+ years old
6   group
7 # 2002/03 nzhs does not have a separate child data
8
9 ac06 <- merge(a06, c06, all=TRUE)
10 ac11 <- merge(a11, c11, all=TRUE)
11 ac12 <- merge(a12, c12, all=TRUE)
12 ac13 <- merge(a13, c13, all=TRUE)
13 ac14 <- merge(a14, c14, all=TRUE)
14
15 # add year variable to all data sets
16 hs02$year <- rep("2002/03")
17 ac06$year <- rep("2006/07")
18 ac11$year <- rep("2011/12")
19 ac12$year <- rep("2012/13")
20 ac13$year <- rep("2013/14")
21 ac14$year <- rep("2014/15")
22
23 # Split them into child data (2–17 years old) and adult data (18+ years old)
24
25 adult <- subset(hs02, age >=18) %>%
26   merge(subset(ac06, age >=18), all=TRUE) %>%
27   merge(subset(ac11, age >=18), all=TRUE) %>%
28   merge(subset(ac12, age >=18), all=TRUE) %>%
29   merge(subset(ac13, age >=18), all=TRUE) %>%
30   merge(subset(ac14, age >=18), all=TRUE)
31
32 child <- subset(ac06, age >=2 & age <18) %>%
33   merge(subset(ac11, age >=2 & age <18), all=TRUE) %>%
34   merge(subset(ac12, age >=2 & age <18), all=TRUE) %>%
35   merge(subset(ac13, age >=2 & age <18), all=TRUE) %>%
36   merge(subset(ac14, age >=2 & age <18), all=TRUE)
37
38 ##### Missing Data analysis #####
```

```

39 ## Missing BMI data
40 # adults
41 adult %>%
42   group_by(year) %>%
43   summarise(n=n(), nmiss=sum(is.na(bmiscale))) %>%
44   mutate(prop.miss = nmiss/n)
45
46 # children
47 child %>%
48   group_by(year) %>%
49   summarise(n=n(), nmiss=sum(is.na(bmic))) %>%
50   mutate(prop.miss = nmiss/n)
51
52
53 ## Missing Household Income
54 # adults
55 adult %>%
56   group_by(year) %>%
57   summarise(n=n(), nmiss=sum(is.na(hhinc))) %>%
58   mutate(prop.miss = nmiss/n)
59
60 # children
61 child %>%
62   group_by(year) %>%
63   summarise(n=n(), nmiss=sum(is.na(hhinc))) %>%
64   mutate(prop.miss = nmiss/n)
65
66
67 ## Missing Education Data
68 # adults
69 adult %>%
70   group_by(year) %>%
71   summarise(n=n(), nmiss=sum(is.na(edu))) %>%
72   mutate(prop.miss = nmiss/n)
73
74 # children
75 child %>%
76   group_by(year) %>%
77   summarise(n=n(), nmiss=sum(is.na(edu))) %>%
78   mutate(prop.miss = nmiss/n)

```



```

79
80
81
82 ##### Histogram, exclude missing BMI data, include relevant variables #####
83
84 # histogram and qq plot
85 normal_plot(adult, "bmiscale")
86 # not normally distributed
87
88 # BMI histogram across ethnicities
89 h_eu = ggplot(subset(adult, eth_count == "European Only"), aes(bmiscale)) +
90   geom_histogram() +
91   labs(title="European Only") +
92   scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))
93
94 h_maori = ggplot(subset(adult, eth_count == "Maori Only"), aes(bmiscale)) +
95   geom_histogram() +
96   labs(title="Maori Only") +
97   scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))
98
99 h_pa = ggplot(subset(adult, eth_count == "Pacific Only"), aes(bmiscale)) +
100   geom_histogram() +
101   labs(title="Pacific Only") +
102   scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))
103
104 h_asian = ggplot(subset(adult, eth_count == "Asian Only"), aes(bmiscale)) +
105   geom_histogram() +
106   labs(title="Asian Only") +
107   scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))
108
109 h_2 = ggplot(subset(adult, eth_count == "2+ Ethnicities (M)"), aes(bmiscale)) +
110   geom_histogram() +
111   labs(title="2+ Ethnicities (M)") +
112   scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))
113
114 h_other = ggplot(subset(adult, eth_count == "Other"), aes(bmiscale)) +
115   geom_histogram() +
116   labs(title="Other") +
117   scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))
118

```

```

119 grid.arrange(h_eu,h_maori,h_pa,h_asian,h_2,h_other , nrow=6, ncol=1)
120
121
122
123 # measure skewness and kurtosis for bmiscale (numeric)
124 skewness(adult$bmiscale , na.rm=TRUE)
125 kurtosis(adult$bmiscale , na.rm=TRUE)
126
127 # Agostino test (H0 = skewness equal to zero)
128 agostino.test(adult$bmiscale , alternative="less") # sample size is too large
129
130 # Anscombe–Glynn test (H0 = kurtosis equal to 3)
131 anscombe.test(adult$bmiscale , alternative="less")
132
133
134 # Subset the Adult data to exclude missing BMI value
135 adult <- subset(adult , !is.na(bmiscale))
136
137
138 # Transform them into tertiles
139 adult$bmi_t <- factor( ntile(adult$bmiscale , 3) ,
140                       labels=c("Bottom Tertile","Mid Tertile","Top Tertile") , ordered=TRUE)
141
142
143 # Identify the BMI range on each tertile
144 adult %>%
145   group_by(bmi_t) %>%
146   summarise_each(funs(min , max) , bmiscale)
147
148
149 ## Subset the data to only include relevant variables
150
151 adult <- subset(adult , select = c("age","bmic","gender","bmiscale" , "UA" , "bmi_t" ,
152                                   "dep" , "hhinc","edu" , "active" , "native" ,
153                                   "sedentary","stair","fruit","veges" ,
154                                   "year","smoke" , "eth_count" , "finalwgt" , "cluster" , "strata" ,
155                                   "haz_drinker_all"))
156
157 # also exclude missing BMI value in the Child data
158 child <- subset(child , !is.na(bmic))

```

```

159
160 child <- subset(child, select=c("age", "bmic", "gender", "bmiscale", "UA",
161                                "dep", "hhinc", "edu", "fruit", "veges",
162                                "year", "softd", "fastf", "eth_count",
163                                "finalwgt", "strata", "cluster"))
164
165
166
167 ##### Bivariate Analysis – Adult #####
168 # All possible combination of explanatory variables with the outcome variable
169 # ANOVA = numeric x categorical; chi-squared = categorical x categorical
170 # We did not use sample weights in the following analyses
171
172 ## Unweighted Stacked plots
173 # Age X bmi_t (BMI tertile)
174 aage = box_notitle(adult, "bmi_t", "age", "BMI Tertile", "Age", min = 15, gap = 5)
175 summary(lm(age ~ bmi_t, data = adult))
176 anova_table(adult$age, adult$bmi_t)
177
178 t = aggregate(age ~ bmi_t, adult, function(x) c(mean = mean(x), sd = sd(x)))
179
180
181 # a function to copy tables to excel
182 write_excel <- function(x, row.names=TRUE, col.names=TRUE, ...) {
183   write_table(x, "clipboard", sep="\t", row.names=row.names, col.names=col.names, ...)
184 }
185 # copied the object "t" and moved them to excel each time
186
187 # Sex
188 asex = stacked_notitle(adult, "bmi_t", "gender", "Sex", "BMI Tertiles")
189 t = table(adult$bmi_t, adult$gender)
190 t = round(prop.table(table(adult$bmi_t, adult$gender), 1), 2)
191
192 # Ethnicity
193 aeth = stacked_notitle(adult, "bmi_t", "eth_count", "Ethnicity", "BMI Tertile")
194 t = table(adult$bmi_t, adult$eth_count)
195 t = prop.table(table(adult$bmi_t, adult$eth_count), 1)
196
197 # Deprivation quintile
198 adep = stacked_notitle(adult, "bmi_t", "dep", "Deprivation Quintile", "BMI Tertile")

```

```

199 t = table(adult$bmi_t, adult$dep)
200 t = prop.table(table(adult$bmi_t, adult$dep), 1)
201
202 # Urban/Rural
203 aua = stacked_notitle(adult, "bmi_t", "UA", "Urban/Rural Area", "BMI Tertile")
204 t = table(adult$bmi_t, adult$UA)
205 t = prop.table(table(adult$bmi_t, adult$UA), 1)
206
207 # Household Income
208 ainc = stacked_notitle(adult, "bmi_t", "hhinc", "Household Income", "BMI Tertile")
209 t = table(adult$bmi_t, adult$hhinc)
210 t = prop.table(table(adult$bmi_t, adult$hhinc), 1)
211
212 # Education
213 aedu = stacked_notitle(adult, "bmi_t", "edu", "Education", "BMI Tertile")
214 t = table(adult$bmi_t, adult$edu)
215 t = prop.table(table(adult$bmi_t, adult$edu), 1)
216
217 # Fruit Consumption
218 afruit = stacked_notitle(adult, "bmi_t", "fruit", "Fruit Guideline", "BMI Tertile")
219 t = table(adult$bmi_t, adult$fruit)
220 t = prop.table(table(adult$bmi_t, adult$fruit), 1)
221
222 # Vegetable Consumption
223 avege = stacked_notitle(adult, "bmi_t", "veges", "Vegetable Guideline", "BMI Tertile")
224 t = table(adult$bmi_t, adult$veges)
225 t = prop.table(table(adult$bmi_t, adult$veges), 1)
226
227
228 # Migration Status
229 amigrat = stacked_notitle(adult, "bmi_t", "native", "Migration Status", "BMI Tertile")
230 t = table(adult$bmi_t, adult$native)
231 t = prop.table(table(adult$bmi_t, adult$native), 1)
232
233 # Difficulty Climbing Stairs
234 astair = stacked_notitle(adult, "bmi_t", "stair", "Difficulty Climbing \nSeveral Flights \nof Stairs", "BMI Tertile")
235 t = table(adult$bmi_t, adult$stair)
236 t = prop.table(table(adult$bmi_t, adult$stair), 1)
237

```

```

238 # Smoking Status
239 asmoke = stacked_notitle(adult, "bmi_t", "smoke", "Smoking Status", "BMI Tertile")
240 t = table(adult$bmi_t, adult$smoke)
241 t = prop.table(table(adult$bmi_t, adult$smoke), 1)
242
243 # Alcohol Problem
244 aalc = stacked_notitle(adult, "bmi_t", "haz_drinker_all", "Drinking Problem", "BMI Tertile")
245 t = table(adult$bmi_t, adult$haz_drinker_all)
246 t = prop.table(table(adult$bmi_t, adult$haz_drinker_all), 1)
247
248 # Physically Active
249 aact = stacked_notitle(adult, "bmi_t", "active", "Physical Activity", "BMI Tertile")
250 t = table(adult$bmi_t, adult$active)
251 t = prop.table(table(adult$bmi_t, adult$active), 1)
252
253 # Sedentary Lifestyle
254 aseden = stacked_notitle(adult, "bmi_t", "sedentary", "Sedentary Lifestyle", "BMI Tertile")
255 t = table(adult$bmi_t, adult$sedentary)
256 t = prop.table(table(adult$bmi_t, adult$sedentary), 1)
257
258
259 # arrange all the charts
260 grid.arrange(aage, asex, aeth, adep, aua, ainc, aedu, afruit, avege, amigrat, astair, asmoke, aalc,
  aact, aseden,
261               ncol = 3, nrow = 5, padding = unit(3, "line"))
262
263
264 ## Chi-Squared tests (unweighted)
265 label = c("Sex", "Ethnicity", "Deprivation Quintile", "Urban/Rural Area", "Household Income", "
  Education", "Fruit Guideline",
266           "Vegetable Guideline", "Migration Status", "Difficulty Climbing Several Flights of Stairs", "
  Smoking Status", "Alcohol Problem",
267           "Physical Activity", "Sedentary Lifestyle")
268
269 table = rbind(chix_table(adult$bmi_t, adult$gender), chix_table(adult$bmi_t, adult$eth_count), chix_
  table(adult$bmi_t, adult$dep),
270         chix_table(adult$bmi_t, adult$UA), chix_table(adult$bmi_t, adult$hhinc), chix_table(adult$bmi_t,
  adult$edu),
271         chix_table(adult$bmi_t, adult$fruit), chix_table(adult$bmi_t, adult$veges), chix_table(adult$bmi
  _t, adult$native),

```

```

272     chix_table(adult$bmi_t, adult$stair), chix_table(adult$bmi_t, adult$smoke), chix_table(adult$bmi
      _t, adult$haz_drinker_all),
273     chix_table(adult$bmi_t, adult$active), chix_table(adult$bmi_t, adult$sedentary))
274
275 biv_table_adult = cbind(label, table)
276
277 # remove temporary object from memory
278 rm(label)
279 rm(table)
280
281
282 ##### Relevel and set reference level – Adult #####
283
284 str(adult)
285
286 # relevel, set reference level
287 adult$eth_count = as.factor(adult$eth_count)
288 adult$eth_count <- relevel(adult$eth_count, ref="European Only")
289 adult$gender <- relevel(adult$gender, ref="Male")
290 adult$dep = relevel(adult$dep, ref=1)
291 adult$hhinc = relevel(adult$hhinc, ref="<=$15,000")
292 adult$hhinc = droplevels(adult$hhinc)
293 adult$edu = relevel(adult$edu, ref="No Qualification")
294 adult$stair = factor(adult$stair, levels = c("No Difficulty", "A Little Difficult", "A Lot Difficult")
      )
295 adult$stair = relevel(adult$stair, ref="No Difficulty")
296 adult$sedentary = as.factor(adult$sedentary)
297 adult$sedentary = relevel(adult$sedentary, ref="Not Sedentary")
298 adult$smoke = relevel(adult$smoke, ref="Non Smoker")
299 adult$veges = droplevels(adult$veges)
300 adult$veges = relevel(adult$veges, ref="No")
301 adult$fruit = relevel(adult$fruit, ref="No")
302 adult$active = relevel(adult$active, ref="Active")
303 adult$year = as.factor(adult$year)
304 adult$year = relevel(adult$year, ref="2014/15")
305 adult$bmic = ordered(adult$bmic)
306 adult$native = droplevels(adult$native)
307 adult$native = relevel(adult$native, ref="Migrant")
308 adult$haz_drinker_all = droplevels(adult$haz_drinker_all)
309

```

```

310
311
312 ##### Survey Design Object – Adult #####
313
314 adult$finalwgt <- as.numeric(adult$finalwgt)
315 summary(adult$finalwgt) # 1 missing value
316
317 # exclude the datum with missing cluster them from the analysis
318 adult <- subset(adult, !is.na(finalwgt))
319
320 # creating a linear contrast
321 contr.lin <- contr.poly(6)[,1]
322
323 # bind the contrast to the data set
324 adult$contr.lin <- contr.lin[ match (adult$year, c("2002/03", "2006/07", "2011/12",
325                                                    "2012/13", "2013/14", "2014/15"))]
326
327 # Create a complex survey design object for the adult data
328 a.design <- svydesign(
329   id = ~cluster,
330   strata = ~interaction(strata, year),
331   data = adult,
332   weights = ~finalwgt,
333   nest = TRUE
334 )
335
336 print(summary(a.design))
337
338
339 ##### BMI Category by Year – Adult #####
340
341
342 ## stacked plot for BMI categories by year (non age-standardised, using survey weights)
343 a.bmi_plot = svyby(~bmic, ~year, svymean, design=a.design, vartype = c('ci', 'se'))
344 a.bmi_plot = data.frame(a.bmi_plot)
345 a.bmi_plot <- gather(a.bmi_plot, bmic, prop, 2:4, factor_key = TRUE)
346
347 a.bmi_plot = a.bmi_plot %>%
348   group_by(year) %>%
349   mutate(pos = cumsum(prop) - 0.5*prop) # calculating position

```

```

350
351 a.bmi_plot$label_p = paste0(sprintf("%.1f", a.bmi_plot$prop*100), "%")
352
353 bmic_adult_year = ggplot(a.bmi_plot, aes(x=year, y=prop, fill=bmic, levels=bmic)) +
354   geom_bar(stat="identity", width = 0.8) +
355   labs(title = "Adult Data", x= "Year", y= "Density") +
356   scale_fill_discrete(guide=FALSE) +
357   theme(axis.title = element_text(size = 12, face="bold"),
358         plot.title = element_text(size = 14, face="bold", hjust=0.5),
359         legend.title = element_text(size = 12, face="bold"),
360         axis.text.x = element_text(size = 11),
361         axis.text.y = element_text(size = 11)) +
362   geom_text(aes(label=label_p), position = position_stack(vjust = 0.5), size=4, colour = "white") +
363   scale_x_discrete(limits=c("2002/03", "2006/07", "2011/12", "2012/13", "2013/14", "2014/15"),
364                   labels=c("'02/03", "'06/07", "'11/12", "'12/13", "'13/14", "'14/15"))
365
366
367 ## age-standardised prevalence
368 adult$agegrp = cut(adult$age, breaks = c(seq(0,85, by = 5), Inf), right = FALSE)
369
370 obes_a = adult %>%
371   group_by(year, agegrp, eth_count, bmic) %>%
372   filter(!is.na(eth_count)) %>%
373   summarise(n=n())
374
375 obes_a = spread(obes_a, bmic, n)
376
377 # replace NA with 0
378 obes_a[is.na(obes_a)] = 0
379
380 obes_a$total = obes_a$Average + obes_a$Overweight + obes_a$Obese
381
382 obes_a$age_aver = (obes_a$Average/obes_a$total)*100
383 obes_a$age_over = (obes_a$Overweight/obes_a$total)*100
384 obes_a$age_obes = (obes_a$Obese/obes_a$total)*100
385
386 # scandinavian pop proportion
387 scand_prop_adult <- c(rep(0.07, 8), 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0.01)
388 # rescale them so that they add up to 100%
389 scand_prop_adult <- scand_prop_adult*(1/((8*0.07)+0.06+0.05+0.04+0.03+0.02+0.01+0.01))

```



```

390
391 obes_a$scand_prop_adult = scand_prop_adult[ match (obes_a$agegrp, c("[15,20)", "[20,25)",
392                                     "[25,30)", "[30,35)", "[35,40)", "[40,45)", "[45,50)",
393                                     "[50,55)", "[55,60)", "[60,65)", "[65,70)", "[70,75)",
394                                     "[75,80)", "[80,85)", "[85,Inf)")))]
395
396 obes_a$aver_rate = obes_a$age_aver*obes_a$scand_prop_adult
397 obes_a$over_rate = obes_a$age_over*obes_a$scand_prop_adult
398 obes_a$obes_rate = obes_a$age_obes*obes_a$scand_prop_adult
399
400 # add them up
401 obes_a1 = aggregate(cbind(aver_rate, over_rate, obes_rate) ~ year+eth_count, FUN = sum, data = obes_a,
402                     na.rm = TRUE, na.action = NULL)
403
404 obes_a1$over_obes = obes_a1$over_rate+obes_a1$obes_rate
405
406 # plot the obesity prevalence by ethnicity
407 aobes_plot = ggplot(obes_a1, aes(x=year, y=obes_rate, shape=eth_count, group=eth_count), stat="
408     identity") +
409     geom_line(size=1) +
410     geom_point(size=2) +
411     theme_light() +
412     scale_x_discrete(limits=c("2002/03", "2006/07", "2011/12", "2012/13", "2013/14", "2014/15")) +
413     scale_y_continuous(limits=c(0,80), breaks=seq(0,100,10)) +
414     labs(title="Adult Data", x = "Year", y = "Percentage", shape="Ethnicity") +
415     theme(axis.title = element_text(size = 14, face="bold"),
416           legend.title = element_text(size = 14, face="bold"),
417           legend.text = element_text(size = 12),
418           axis.text.x = element_text(size = 12),
419           axis.text.y = element_text(size = 12),
420           legend.position="none",
421           plot.title = element_text(size = 16, face="bold", hjust=0.5))
422
423 # plot the over/obes prev
424 aovob_plot = ggplot(obes_a1, aes(x=year, y=over_rate+obes_rate, shape=eth_count, group=eth_count),
425                     stat="identity") +
426     geom_line(size=1) +
427     geom_point(size=2) +
428     theme_light() +

```

```

428 scale_x_discrete(limits=c("2002/03", "2006/07", "2011/12", "2012/13", "2013/14", "2014/15")) +
429 scale_y_continuous(limits=c(20,100), breaks=seq(0,100,10)) +
430 labs(title="Adult Data", x = "Year", y = "Percentage", shape="Ethnicity") +
431 theme(axis.title = element_text(size = 14, face="bold"),
432        legend.title = element_text(size = 14, face="bold"),
433        legend.text = element_text(size = 12),
434        axis.text.x = element_text(size = 12),
435        axis.text.y = element_text(size = 12),
436        legend.position="none",
437        plot.title = element_text(size = 16, face="bold", hjust=0.5))
438
439
440
441
442 ##### Descriptive analysis across survey periods – Adult #####
443
444 # total sample size on each year
445 adult %>%
446   group_by(year) %>%
447   summarise(row = length(age))
448
449 # age summary
450 svyquantile(~age, subset(a.design, year == "2002/03"), c(0,.25,.5,.75,1), ci=TRUE)
451 svyquantile(~age, subset(a.design, year == "2006/07"), c(0,.25,.5,.75,1), ci=TRUE)
452 svyquantile(~age, subset(a.design, year == "2011/12"), c(0,.25,.5,.75,1), ci=TRUE)
453 svyquantile(~age, subset(a.design, year == "2012/13"), c(0,.25,.5,.75,1), ci=TRUE)
454 svyquantile(~age, subset(a.design, year == "2013/14"), c(0,.25,.5,.75,1), ci=TRUE)
455 svyquantile(~age, subset(a.design, year == "2014/15"), c(0,.25,.5,.75,1), ci=TRUE)
456
457 # bmi summary
458 svyquantile(~bmiscale, subset(a.design, year == "2002/03"), c(0,.25,.5,.75,1), ci=TRUE)
459 svyquantile(~bmiscale, subset(a.design, year == "2006/07"), c(0,.25,.5,.75,1), ci=TRUE)
460 svyquantile(~bmiscale, subset(a.design, year == "2011/12"), c(0,.25,.5,.75,1), ci=TRUE)
461 svyquantile(~bmiscale, subset(a.design, year == "2012/13"), c(0,.25,.5,.75,1), ci=TRUE)
462 svyquantile(~bmiscale, subset(a.design, year == "2013/14"), c(0,.25,.5,.75,1), ci=TRUE)
463 svyquantile(~bmiscale, subset(a.design, year == "2014/15"), c(0,.25,.5,.75,1), ci=TRUE)
464
465
466 # categorical variables
467 svyby(~bmic, ~year, svymean, design=a.design)

```

```

468 svyby(~gender, ~year, svymean, design=a.design)
469 svyby(~eth_count, ~year, svymean, design=a.design)
470 svyby(~dep, ~year, svymean, design=a.design)
471 svyby(~UA, ~year, svymean, design=a.design)
472 svyby(~hhinc, ~year, svymean, design=a.design)
473 svyby(~edu, ~year, svymean, design=a.design)
474 svyby(~fruit, ~year, svymean, design=a.design)
475 svyby(~veges, ~year, svymean, design=a.design)
476 svyby(~native, ~year, svymean, design=a.design)
477 svyby(~stair, ~year, svymean, design=a.design)
478 svyby(~smoke, ~year, svymean, design=a.design)
479 svyby(~haz_drinker_all, ~year, svymean, design=a.design)
480 svyby(~active, svymean, ~year, design=a.design)
481 svyby(~sedentary, ~year, svymean, design=a.design)
482
483
484
485 ##### Trends over the survey periods – Adult #####
486 # reference level = 2014/15
487 summy(svyglm(age ~ year, a.design))
488 summy(svyglm(l(bmic == "Obese") ~ year, a.design))
489 summy(svyglm(l(gender == "Male") ~ year, a.design))
490 summy(svyglm(l(eth_count == "European Only") ~ year, a.design))
491 summy(svyglm(l(eth_count == "Other") ~ year, a.design))
492 summy(svyglm(l(UA == "Urban") ~ year, a.design))
493 summy(svyglm(l(fruit == "Yes") ~ year, a.design))
494 summy(svyglm(l(veges == "Yes") ~ year, a.design))
495 summy(svyglm(l(active == "Active") ~ year, a.design))
496 summy(svyglm(l(sedentary == "Sedentary") ~ year, a.design))
497
498
499 ##### Forward Step Proportional Odds Logistic Regression – Adult #####
500 # Add one covariate at a time, add the strongest covariate on each step, and remove any non
    significant variable (p < 0.1)
501
502 ## 1st step
503 # add deprivation and linear changes, variables of interest
504 olr1_dep = svyolr(bmi_t ~ dep, a.design)
505 summy(olr1_dep)
506

```

```

507 olr1_dep.lin = svyolr(bmi_t ~ dep + contr.lin , a.design)
508 summy(olr1_dep.lin)
509
510 # add other explanatory variables and try to find the strongest effect to add into the next step
511 olr1_inc = svyolr(bmi_t ~ hhinc + dep + contr.lin , a.design)
512 olr1_vege = svyolr(bmi_t ~ veges + dep + contr.lin , a.design)
513 olr1_age = svyolr(bmi_t ~ age + dep + contr.lin , a.design)
514 olr1_sex = svyolr(bmi_t ~ gender + dep + contr.lin , a.design)
515 olr1_eth = svyolr(bmi_t ~ eth_count + dep + contr.lin , a.design)
516 olr1_edu = svyolr(bmi_t ~ edu + dep + contr.lin , a.design)
517 olr1_seden = svyolr(bmi_t ~ sedentary + dep + contr.lin , a.design)
518 olr1_act = svyolr(bmi_t ~ active + dep + contr.lin , a.design)
519 olr1_stair = svyolr(bmi_t ~ stair + dep + contr.lin , a.design)
520 olr1_smoke = svyolr(bmi_t ~ smoke + dep + contr.lin , a.design)
521 olr1_fruit = svyolr(bmi_t ~ fruit + dep + contr.lin , a.design)
522 olr1_native = svyolr(bmi_t ~ native + dep + contr.lin , a.design)
523 olr1_alc = svyolr(bmi_t ~ haz_drinker_all + dep + contr.lin , a.design)
524 olr1_UA = svyolr(bmi_t ~ UA + dep + contr.lin , a.design)
525
526 summy(olr1_inc)
527 summy(olr1_vege)
528 summy(olr1_age)
529 summy(olr1_sex)
530 summy(olr1_eth)
531 summy(olr1_edu)
532 summy(olr1_seden)
533 summy(olr1_act)
534 summy(olr1_stair)
535 summy(olr1_smoke)
536 summy(olr1_fruit)
537 summy(olr1_native)
538 summy(olr1_alc)
539 summy(olr1_UA)
540 # alcohol problem and adherence to vegetable guideline were insignificant
541 # Ethnicity had the strongest effect
542
543
544 ## 2nd step (bmi_t ~ ... + dep + contr.lin , a.design)
545 olr2_1 = svyolr(bmi_t ~ age + eth_count + dep + contr.lin , a.design)
546 olr2_2 = svyolr(bmi_t ~ gender + eth_count + dep + contr.lin , a.design)

```

```

547 olr2_3 = svyolr(bmi_t ~ edu + eth_count + dep + contr.lin , a.design)
548 olr2_4 = svyolr(bmi_t ~ sedentary + eth_count + dep + contr.lin , a.design)
549 olr2_5 = svyolr(bmi_t ~ stair + eth_count + dep + contr.lin , a.design)
550 olr2_6 = svyolr(bmi_t ~ smoke + eth_count + dep + contr.lin , a.design)
551 olr2_7 = svyolr(bmi_t ~ age + hhinc + dep + contr.lin , a.design)
552 olr2_8 = svyolr(bmi_t ~ active + eth_count + dep + contr.lin , a.design)
553 olr2_9 = svyolr(bmi_t ~ native + eth_count + dep + contr.lin , a.design)
554 olr2_10 = svyolr(bmi_t ~ UA + eth_count + dep + contr.lin , a.design)
555 olr2_11 = svyolr(bmi_t ~ hhinc + eth_count + dep + contr.lin , a.design)
556 olr2_12 = svyolr(bmi_t ~ fruit + eth_count + dep + contr.lin , a.design)
557
558 summy(olr2_1)
559 summy(olr2_2)
560 summy(olr2_3)
561 summy(olr2_4)
562 summy(olr2_5)
563 summy(olr2_6)
564 summy(olr2_7)
565 summy(olr2_8)
566 summy(olr2_9)
567 summy(olr2_10)
568 summy(olr2_11)
569 summy(olr2_12)
570 # adherence to fruit guideline was insignificant
571 # Age had the strongest effect
572
573
574 ## 3rd step
575 olr3_1 = svyolr(bmi_t ~ gender + age + eth_count + dep + contr.lin , a.design)
576 olr3_2 = svyolr(bmi_t ~ smoke + age + eth_count + dep + contr.lin , a.design)
577 olr3_3 = svyolr(bmi_t ~ sedentary + age + eth_count + dep + contr.lin , a.design)
578 olr3_4 = svyolr(bmi_t ~ edu + age + eth_count + dep + contr.lin , a.design)
579 olr3_5 = svyolr(bmi_t ~ stair + age + eth_count + dep + contr.lin , a.design)
580 olr3_6 = svyolr(bmi_t ~ active + age + eth_count + dep + contr.lin , a.design)
581 olr3_7 = svyolr(bmi_t ~ native + age + eth_count + dep + contr.lin , a.design)
582 olr3_8 = svyolr(bmi_t ~ UA + age + eth_count + dep + contr.lin , a.design)
583 olr3_9 = svyolr(bmi_t ~ hhinc + age + eth_count + dep + contr.lin , a.design)
584 olr3_10 = svyolr(bmi_t ~ fruit + age + eth_count + dep + contr.lin , a.design)
585
586 summy(olr3_1)

```

```

587 summy(olr3_2)
588 summy(olr3_3)
589 summy(olr3_4)
590 summy(olr3_5)
591 summy(olr3_6)
592 summy(olr3_7)
593 summy(olr3_8)
594 summy(olr3_9)
595 summy(olr3_10)
596 # Difficulty climbing several steps of stairs had the strongest effect
597
598
599 ## 4th step
600 olr4_1 = svyolr(bmi_t ~ gender + stair + age + eth_count + dep + contr.lin , a.design)
601 olr4_2 = svyolr(bmi_t ~ edu + stair + age + eth_count + dep + contr.lin , a.design)
602 olr4_3 = svyolr(bmi_t ~ sedentary + stair + age + eth_count + dep + contr.lin , a.design)
603 olr4_4 = svyolr(bmi_t ~ smoke + stair + age + eth_count + dep + contr.lin , a.design)
604 olr4_5 = svyolr(bmi_t ~ active + stair + age + eth_count + dep + contr.lin , a.design)
605 olr4_6 = svyolr(bmi_t ~ native + stair + age + eth_count + dep + contr.lin , a.design)
606 olr4_7 = svyolr(bmi_t ~ UA + stair + age + eth_count + dep + contr.lin , a.design)
607 olr4_8 = svyolr(bmi_t ~ hhinc + stair + age + eth_count + dep + contr.lin , a.design)
608
609 summy(olr4_1)
610 summy(olr4_2)
611 summy(olr4_3)
612 summy(olr4_4)
613 summy(olr4_5)
614 summy(olr4_6)
615 summy(olr4_7)
616 summy(olr4_8)
617
618 # Sedentary lifestyle lost its significance
619 # Gender had the strongest effect
620
621
622 ## 5th step
623 olr5_1 = svyolr(bmi_t ~ edu + gender + stair + age + eth_count + dep + contr.lin , a.design)
624 olr5_2 = svyolr(bmi_t ~ smoke + gender + stair + age + eth_count + dep + contr.lin , a.design)
625 olr5_3 = svyolr(bmi_t ~ active + gender + stair + age + eth_count + dep + contr.lin , a.design)
626 olr5_4 = svyolr(bmi_t ~ native + gender + stair + age + eth_count + dep + contr.lin , a.design)

```

```

627 olr5_5 = svyolr(bmi_t ~ UA + gender + stair + age + eth_count + dep + contr.lin , a.design)
628 olr5_6 = svyolr(bmi_t ~ hhinc + gender + stair + age + eth_count + dep + contr.lin , a.design)
629
630 summy(olr5_1)
631 summy(olr5_2)
632 summy(olr5_3)
633 summy(olr5_4)
634 summy(olr5_5)
635 summy(olr5_6)
636 # Ex-smoker had the strongest effect
637
638
639 ## 6th step
640 olr6_1 = svyolr(bmi_t ~ edu + smoke + gender + stair + age + eth_count + dep + contr.lin , a.design)
641 olr6_2 = svyolr(bmi_t ~ active + smoke + gender + stair + age + eth_count + dep + contr.lin , a.
        design)
642 olr6_3 = svyolr(bmi_t ~ native + smoke + gender + stair + age + eth_count + dep + contr.lin , a.
        design)
643 olr6_4 = svyolr(bmi_t ~ UA + smoke + gender + stair + age + eth_count + dep + contr.lin , a.design)
644 olr6_5 = svyolr(bmi_t ~ hhinc + smoke + gender + stair + age + eth_count + dep + contr.lin , a.design
        )
645
646 summy(olr6_1)
647 summy(olr6_2)
648 summy(olr6_3)
649 summy(olr6_4)
650 summy(olr6_5)
651 # Urban/rural area lose its significance
652 # Household income had the strongest effect
653
654
655 ## 7th step
656 olr7_1 = svyolr(bmi_t ~ edu + hhinc + smoke + gender + stair + age + eth_count + dep + contr.lin , a.
        design)
657 olr7_2 = svyolr(bmi_t ~ native + hhinc + smoke + gender + stair + age + eth_count + dep + contr.lin ,
        a.design)
658 olr7_3 = svyolr(bmi_t ~ active + hhinc + smoke + gender + stair + age + eth_count + dep + contr.lin ,
        a.design)
659
660 summy(olr7_1)

```

```

661 summy(olr7_2)
662 summy(olr7_3)
663 # Education had the strongest effect
664
665
666 ## 8th step
667 olr8_1 = svyolr(bmi_t ~ active + edu + hhinc + smoke + gender +
668               stair + age + eth_count + dep + contr.lin , a.design)
669 olr8_2 = svyolr(bmi_t ~ native + edu + hhinc + smoke + gender +
670               stair + age + eth_count + dep + contr.lin , a.design)
671
672 summy(olr8_1)
673 summy(olr8_2)
674 # Physical activity had the strongest effect
675
676
677 ## 9th step
678 olr9 = svyolr(bmi_t ~ native + active + edu + hhinc + smoke + gender +
679               stair + age + eth_count + dep + contr.lin , a.design)
680
681 summy(olr9 , round=4)
682 # Migration status remained significant
683 # olr9 is the final model
684
685 # calculating one-tailed z-test for the deprivation quintiles
686 # Alt. Hypothesis = The likelihood of being in the higher BMI tertiles is higher for people living
687 # in areas with higher deprivation
688 pnorm(coefficients(summary(olr9))[19,3], lower.tail=FALSE) # quintile 2
689 pnorm(coefficients(summary(olr9))[20,3], lower.tail=FALSE) # quintile 3
690 pnorm(coefficients(summary(olr9))[21,3], lower.tail=FALSE) # quintile 4
691 pnorm(coefficients(summary(olr9))[22,3], lower.tail=FALSE) # quintile 5
692
693
694 ##### Proportional Odds Assumption – Adult #####
695
696 # Binomial regression using contrast
697 # ex1 assessing the odds of being in the Mid or Top tertile vs. the Bottom tertile
698 # ex2 assessing the odds of being in the Top tertile vs. Bottom or Mid tertile
699 ex1 <- svyglm(l(bmi_t>"Bottom Tertile") ~ dep + contr.lin + eth_count + age + stair +

```



```

700         gender + smoke + hhinc + edu + active + native, a.design, family = "binomial")
701 ex2 <- svyglm(l(bmi_t>"Mid Tertile") ~ dep + contr.lin + eth_count + age + stair +
702         gender + smoke + hhinc + edu + active + native, a.design, family = "binomial")
703
704 # plot them
705 plot(coef(ex1)[-1], xlim=c(0,30), ylim=c(-1,2), type="l", ylab="Beta", main="Comparison of Betas for
       \nProportional Odds Assumption")
706 lines(coef(ex2)[-1], col=2)
707
708 # Standard Errors and Coefficients
709 rbind(SE(ex1),SE(ex2))[, -1]
710 rbind(coef(ex1),coef(ex2))[, -1]
711
712 # odds ratios
713 rbind(exp(coef(ex1)),exp(coef(ex2)))[-1]
714
715 # 2.5% CL odds ratio
716 rbind(exp(coef(ex1)-SE(ex1)), exp(coef(ex2)-SE(ex2)))[-1]
717
718 # 97.5% CL odds ratio
719 rbind(exp(coef(ex1)+SE(ex1)), exp(coef(ex2)+SE(ex2)))[-1]
720
721
722 ##### Bivariate Analysis – Child #####
723 # All possible combination of explanatory variables with the outcome variable
724 # ANOVA = numeric x categorical; chi-squared = categorical x categorical
725 # Excluding missing BMI value
726
727 ## Unweigthed Stacked plots
728 # Age X bmic (bmi category)
729 cage = box_notitle(child, "bmic", "age", "BMI Category", "Age", min=2, max = 18, bins = 2, gap =
       1.5)
730 summary(lm(age ~ bmic, data = child))
731 anova_table(child$age, child$bmic)
732
733 t = aggregate(age ~ bmic, child, function(x) c(mean=mean(x),sd=sd(x)))
734
735 # Sex
736 csex = stacked_notitle(child, "bmic", "gender", "Sex", "BMI Category")
737 t = table(child$bmic, child$gender)

```

```

738 t = prop.table(table(child$bmic, child$gender), 1)
739
740 # Ethnicity
741 ceth = stacked_notitle(child, "bmic", "eth_count", "Ethnicity", "BMI Category")
742 t = table(child$bmic, child$eth_count)
743 t = prop.table(table(child$bmic, child$eth_count), 1)
744
745 # Deprivation Quintile
746 cdep = stacked_notitle(child, "bmic", "dep", "Deprivation Quintile", "BMI Category")
747 t = table(child$bmic, child$dep)
748 t = prop.table(table(child$bmic, child$dep), 1)
749
750 # Urban/Rural
751 cua = stacked_notitle(child, "bmic", "UA", "Urban/Rural Area", "BMI Category")
752 t = table(child$bmic, child$UA)
753 t = prop.table(table(child$bmic, child$UA), 1)
754
755 # Household Income
756 cinc = stacked_notitle(child, "bmic", "hhinc", "Household Income", "BMI Category")
757 t = table(child$bmic, child$hhinc)
758 t = prop.table(table(child$bmic, child$hhinc), 1)
759
760 # Education
761 cedu = stacked_notitle(child, "bmic", "edu", "Education", "BMI Category")
762 t = table(child$bmic, child$edu)
763 t = prop.table(table(child$bmic, child$edu), 1)
764
765 # Fruit Guideline
766 cfruit = stacked_notitle(child, "bmic", "fruit", "Fruit Guideline", "BMI Category")
767 t = table(child$bmic, child$fruit)
768 t = prop.table(table(child$bmic, child$fruit), 1)
769
770 # Vegetable Guideline
771 cvege = stacked_notitle(child, "bmic", "veges", "Vegetable Guideline", "BMI Category")
772 t = table(child$bmic, child$veges)
773 t = prop.table(table(child$bmic, child$veges), 1)
774
775 # Soft Drink Consumption
776 csoft = stacked_notitle(child, "bmic", "softd", "Soft Drink Consumption", "BMI Category")
777 t = table(child$bmic, child$softd)

```

```

778 t = prop.table(table(child$bmic, child$softd), 1)
779
780 # Fast Food Consumption
781 cfast = stacked_notitle(child, "bmic", "fastf", "Fast Food Consumption", "BMI Category")
782 t = table(child$bmic, child$fastf)
783 t = prop.table(table(child$bmic, child$fastf), 1)
784
785
786 # arrange the charts
787 grid.arrange(cage, csex, ceth, cdep, cua, cinc, cedu, cfruit, cvege, csoft, cfast,
788             ncol = 3, nrow=4, padding=unit(3, "line"))
789
790
791
792 ## Chi-Squared tests (unweighted)
793 label = c("Sex", "Ethnicity", "Deprivation Quintile", "Urban/Rural Area", "Household Income", "
794           Education", "Fruit Guideline",
795           "Vegetable Guideline", "Soft Drink Consumption", "Fast Food Consmpion")
796
797 table = rbind(chix_table(child$bmic, child$gender), chix_table(child$bmic, child$eth_count), chix_
798               table(child$bmic, child$dep),
799               chix_table(child$bmic, child$UA), chix_table(child$bmic, child$hhinc), chix_table(child$bmic,
800               child$edu),
801               chix_table(child$bmic, child$fruit), chix_table(child$bmic, child$veges), chix_table(child$bmic,
802               child$softd),
803               chix_table(child$bmic, child$fastf))
804
805 biv_table_child = cbind(label, table)
806
807
808 # remove unused object from memory
809 rm(label)
810 rm(table)
811
812
813 ##### Relevel and set reference level – Child #####
814
815 child$veges = droplevels(child$veges)
816 child$eth_count = relevel(child$eth_count, ref="European Only")
817 child$hhinc = relevel(child$hhinc, ref=">$70,000")
818 child$bmic = ordered(child$bmic)

```

```

814 child$year = as.factor(child$year)
815 child$year = relevel(child$year, ref="2014/15")
816 child$softd = relevel(child$softd, ref="0/week")
817 child$fastf = relevel(child$fastf, ref="0/week")
818
819 child$soft_order = factor(child$softd, levels=c("0/week", "1/week", "2-3/week", "4+/week"), ordered=
      TRUE)
820 child$fast_order = factor(child$fastf, levels=c("0/week", "1/week", "2-3/week", "4+/week"), ordered=
      TRUE)
821
822
823 ##### Survey Design Object – Child #####
824
825 # missing sample weights
826 child$finalwgt = as.numeric(child$finalwgt)
827 sum(is.na(child$finalwgt)) # 4 NA
828
829 # remove them from analysis
830 child = subset(child, !is.na(finalwgt))
831
832 # creating a linear contrast
833 contr.lin <- contr.poly(5)[,1]
834
835 # bind the contrast to the data set
836 child$contr.lin <- contr.lin[ match (child$year, c("2006/07", "2011/12",
837                                                    "2012/13", "2013/14", "2014/15"))]
838
839 rm(contr.lin)
840
841 # complex survey design
842 c.design <- svydesign(
843   id = ~cluster ,
844   strata = ~interaction(strata , year) ,
845   data = child ,
846   weights = ~finalwgt ,
847   nest = TRUE
848 )
849
850 summary(c.design)
851

```

```

852
853 ##### Obesity Prevalence by Year – Child #####
854
855
856 ## BMI categories by year (non-age standardised, using survey weights)
857
858 bmic_year <- svyby(~bmic, ~year, svymean, design= c.design)
859 bmic_year <- gather(data.frame(bmic_year), bmic, prop, 2:4, factor_key=TRUE)
860
861 bmic_year = bmic_year %>%
862   group_by(year) %>%
863   mutate(pos = cumsum(prop) - 0.5*prop) # calculating position
864
865 bmic_year$label_p = paste0(sprintf("%.1f", bmic_year$prop*100), "%")
866
867 # bar plot
868 bmi_child_year = ggplot(bmic_year, aes(x=year, y=prop, fill=bmic, levels=bmic)) +
869   geom_bar(stat="identity", width = 0.8) +
870   labs(title = "Child Data", x= "Year", y= "Density") +
871   scale_fill_discrete(name = "BMI Category", labels = c("Average", "Overweight", "Obese")) +
872   theme(axis.title = element_text(size = 12, face="bold"),
873         plot.title = element_text(size = 14, face="bold", hjust=0.5),
874         legend.title = element_text(size = 12, face="bold"),
875         axis.text.x = element_text(size = 11),
876         axis.text.y = element_text(size = 11)) +
877   geom_text(aes(label=label_p), position = position_stack(vjust = 0.5), size=4, colour = "white") +
878   scale_x_discrete(limits=c("2006/07", "2011/12", "2012/13", "2013/14", "2014/15"),
879                     labels=c("'06/07", "'11/12", "'12/13", "'13/14", "'14/15"))
880
881
882 plot_grid(bmic_adult_year, bmi_child_year, align = "h", rel_widths=c(0.8,1))
883
884
885 ## calculate bmi categories by ethnicity and year ()
886 # create age group variable
887 child$agegrp = cut(child$age, breaks = c(seq(0,85, by = 5), Inf), right = FALSE)
888
889
890 # calculate rates per 100 population (or proportion of BMI categories by year, ethnicity, age group)
891 obes_c = child %>%

```

```

892 group_by(year, agegrp, eth_count, bmic) %>%
893   filter(!is.na(eth_count)) %>%
894   summarise(n=n()) %>%
895   mutate(prop = n/sum(n))
896
897
898 # add scandinavian population proportion and rescale them so that they add up to 100%
899 scand_prop_child = c(0.08,0.07,0.07,0.07)
900 scand_prop_child = scand_prop_child*(1/(0.08 + 0.07 + 0.07 + 0.07))
901
902 # match them up
903 obes_c$scand_prop_child = scand_prop_child[ match (obes_c$agegrp, c("[0,5)", "[5,10)", "[10,15)",
904   "[15,20)"))]]
905
906 # calculate age-specific rate
907 obes_c$age_r = obes_c$prop*obes_c$scand_prop_child
908
909
910 cage_std_plot = aggregate(age_r ~ year+eth_count+bmic, FUN=sum, data=obes_c, na.rm=TRUE)
911
912 cage_std_plot = spread(cage_std_plot, bmic, age_r)
913
914 cage_std_plot[,3:5] = sapply(cage_std_plot[,3:5], function(x) x*100)
915
916 cage_std_plot$over_obes = cage_std_plot$Overweight+cage_std_plot$Obese
917
918
919 ## Obesity rate by year and ethncity
920 cobes_plot = ggplot(cage_std_plot, aes(x=year, y=Obese, shape=eth_count, group=eth_count), stat="
  identity") +
921   geom_line(size=1) +
922   geom_point(size=2) +
923   theme_light() +
924   scale_x_discrete(limits=c("2006/07", "2011/12", "2012/13", "2013/14", "2014/15")) +
925   scale_y_continuous(limits=c(0,80), breaks=seq(0,100,10)) +
926   labs(title="Child Data", x = "Year", y = "Percentage", shape="Ethnicity") +
927   theme(axis.title = element_text(size = 14, face="bold"),
928         legend.title = element_text(size = 14, face="bold"),
929         legend.text = element_text(size = 12),

```

```

930     axis.text.x = element_text(size = 12),
931     axis.text.y = element_text(size = 12),
932     plot.title = element_text(size = 16, face="bold", hjust=0.5))
933
934 plot_grid(aobes_plot, cobes_plot, align = "h", rel_widths=c(0.8,1))
935
936
937
938 ## Overweight and obesity rate by year and ethnicity
939 covob_plot = ggplot(cage_std_plot, aes(x=year, y=Overweight+Obese, shape=eth_count, group=eth_count)
940   , stat="identity") +
941   geom_line(size=1) +
942   geom_point(size=2) +
943   theme_light() +
944   scale_x_discrete(limits=c("2006/07", "2011/12", "2012/13", "2013/14", "2014/15")) +
945   scale_y_continuous(limits=c(20,100), breaks=seq(0,100,10)) +
946   labs(title="Child Data", x = "Year", y = "Percentage", shape="Ethnicity") +
947   theme(axis.title = element_text(size = 14, face="bold"),
948     legend.title = element_text(size = 14, face="bold"),
949     legend.text = element_text(size = 12),
950     axis.text.x = element_text(size = 12),
951     axis.text.y = element_text(size = 12),
952     plot.title = element_text(size = 16, face="bold", hjust=0.5))
953
954 plot_grid(aovob_plot, covob_plot, align = "h", rel_widths=c(0.8,1))
955
956
957 ##### Descriptive Analysis across survey periods – Child #####
958
959 # total sample size on each year
960 child %>%
961   group_by(year) %>%
962   summarise(row = length(age))
963
964 # age summary
965 svyquantile(~age, subset(c.design, year == "2006/07"), c(0,.25,.5,.75,1), ci=TRUE)
966 svyquantile(~age, subset(c.design, year == "2011/12"), c(0,.25,.5,.75,1), ci=TRUE)
967 svyquantile(~age, subset(c.design, year == "2012/13"), c(0,.25,.5,.75,1), ci=TRUE)
968 svyquantile(~age, subset(c.design, year == "2013/14"), c(0,.25,.5,.75,1), ci=TRUE)

```

```

969 svyquantile(~age, subset(c.design, year == "2014/15"), c(0,.25,.5,.75,1), ci=TRUE)
970
971
972 # categorical variables
973 svyby(~bmic, ~year, svymean, design=c.design)
974 svyby(~gender, ~year, svymean, design=c.design)
975 svyby(~eth_count, ~year, svymean, design=c.design)
976 svyby(~dep, ~year, ~year, svymean, design=c.design)
977 svyby(~UA, ~year, svymean, design=c.design)
978 svyby(~hhinc, ~year, svymean, design=c.design)
979 svyby(~edu, ~year, svymean, design=c.design)
980 svyby(~fruit, ~year, svymean, design=c.design)
981 svyby(~veges, ~year, svymean, design=c.design)
982 svyby(~softd, ~year, svymean, design=c.design)
983 svyby(~fastf, ~year, svymean, design=c.design)
984
985
986 ##### Trends across survey periods – Child #####
987 # reference level = 2014/15
988 summy(svyglm(l(bmic == "Obese") ~ year, c.design))
989 c.design$variables$year = relevel(c.design$variables$year, ref= "2011/12")
990 summy(svyglm(l(bmic == "Obese") ~ year, c.design)) # no diff 2011/12 vs 2006/07
991 c.design$variables$year = relevel(c.design$variables$year, ref= "2012/13")
992 summy(svyglm(l(bmic == "Obese") ~ year, c.design)) # sig. diff: 2012/13 vs 2006/07
993
994 # reset the reference group to "2014/15"
995 c.design$variables$year = factor(c.design$variables$year, levels=c("2002/03",
996                             "2006/07", "2011/12", "2012/13", "2013/14", "2014/15"))
997 c.design$variables$year = relevel(c.design$variables$year, ref= "2014/15")
998
999
1000 summy(svyglm(l(eth_count == "European Only") ~ year, c.design))
1001 summy(svyglm(l(eth_count == "Other") ~ year, c.design))
1002 summy(svyglm(l(UA == "Urban") ~ year, c.design))
1003 summy(svyglm(l(edu == "Tertiary") ~ year, c.design))
1004 summy(svyglm(l(soft_order > "1/week") ~ year, c.design))
1005 summy(svyglm(l(fast_order > "1/week") ~ year, c.design))
1006
1007
1008 ##### Forward Step Proportional Odds Logistic Regression – Child #####

```



```

1009 # Add one covariate at a time, add the strongest covariate on each step, and remove any non
      significant variable (p < 0.1)
1010
1011 ## 1st step
1012 # add deprivation quintile and linear contrasts, variables of interest
1013 colr1_1 = svyolr(bmic ~ dep + contr.lin, c.design)
1014
1015 summy(colr1_1)
1016
1017
1018 ## 2nd step
1019 colr2_0 = svyolr(bmic ~ gender + dep + contr.lin, c.design)
1020 colr2_1 = svyolr(bmic ~ age + dep + contr.lin, c.design)
1021 colr2_2 = svyolr(bmic ~ eth_count + dep + contr.lin, c.design)
1022 colr2_3 = svyolr(bmic ~ hhinc + dep + contr.lin, c.design)
1023 colr2_4 = svyolr(bmic ~ edu + dep + contr.lin, c.design)
1024 colr2_5 = svyolr(bmic ~ fruit + dep + contr.lin, c.design)
1025 colr2_6 = svyolr(bmic ~ veges + dep + contr.lin, c.design)
1026 colr2_7 = svyolr(bmic ~ softd + dep + contr.lin, c.design)
1027 colr2_8 = svyolr(bmic ~ fastf + dep + contr.lin, c.design)
1028 colr2_9 = svyolr(bmic ~ UA + dep + contr.lin, c.design)
1029
1030 summy(colr2_0)
1031 summy(colr2_1)
1032 summy(colr2_2)
1033 summy(colr2_3)
1034 summy(colr2_4)
1035 summy(colr2_5)
1036 summy(colr2_6)
1037 summy(colr2_7)
1038 summy(colr2_8)
1039 summy(colr2_9)
1040 # fruit, veges and urban/rural area were not significant
1041 # Ethnicity had the strongest effect
1042
1043
1044 ## 3rd step
1045 colr3_1 = svyolr(bmic ~ age + eth_count + dep + contr.lin, c.design)
1046 colr3_2 = svyolr(bmic ~ hhinc + eth_count + dep + contr.lin, c.design)
1047 colr3_3 = svyolr(bmic ~ edu + eth_count + dep + contr.lin, c.design)

```

```

1048 colr3_4 = svyolr(bmic ~ softd + eth_count + dep + contr.lin , c.design)
1049 colr3_5 = svyolr(bmic ~ fastf + eth_count + dep + contr.lin , c.design)
1050 colr3_6 = svyolr(bmic ~ gender + eth_count + dep + contr.lin , c.design)
1051
1052 summy(colr3_1)
1053 summy(colr3_2)
1054 summy(colr3_3)
1055 summy(colr3_4)
1056 summy(colr3_5)
1057 summy(colr3_6)
1058 # household income lost became insignificant
1059 # Age had the strongest effect
1060
1061
1062 ## 4th step
1063 colr4_1 = svyolr(bmic ~ edu + age + eth_count + dep + contr.lin , c.design)
1064 colr4_2 = svyolr(bmic ~ softd + age + eth_count + dep + contr.lin , c.design)
1065 colr4_3 = svyolr(bmic ~ fastf + age + eth_count + dep + contr.lin , c.design)
1066 colr4_4 = svyolr(bmic ~ gender + age + eth_count + dep + contr.lin , c.design)
1067
1068 summy(colr4_4)
1069 summy(colr4_1)
1070 summy(colr4_2)
1071 summy(colr4_3)
1072 # Education had the strongest effect
1073
1074
1075 ## 5th step
1076 colr5_1 = svyolr(bmic ~ softd + edu + age + eth_count + dep + contr.lin , c.design)
1077 colr5_2 = svyolr(bmic ~ fastf + edu + age + eth_count + dep + contr.lin , c.design)
1078 colr5_3 = svyolr(bmic ~ gender + edu + age + eth_count + dep + contr.lin , c.design)
1079
1080 summy(colr5_1)
1081 summy(colr5_2)
1082 summy(colr5_3)
1083 # soft drink consumption had the strongest effect
1084
1085
1086 ## 6th step
1087 colr6_1 = svyolr(bmic ~ fastf + softd + edu + age + eth_count + dep + contr.lin , c.design)

```

```

1088 colr6_2 = svyolr(bmic ~ gender + softd + edu + age + eth_count + dep + contr.lin , c.design)
1089
1090 summy(colr6_1)
1091 summy(colr6_2)
1092 # fast food consumption lost was not significant , maybe due to colinearity with soft drink
      consumption
1093
1094
1095 ## Model Fit Tests
1096 ## Model 1 (0.05 rule): bm1c ~ softd + edu + age + eth_count + dep + contr.lin , c.design
1097 ## Model 2 (0.1 rule): bm1c ~ gender + softd + edu + age + eth_count + dep + contr.lin , c.design
1098 ## Model 3 (0.2 rule): bm1c ~ fastf + gender + softd + edu + age + eth_count + dep + contr.lin , c.
      design
1099
1100 # does adding gender into model 1 is better?
1101 model1 = svyolr(bmic ~ softd + edu + age + eth_count + dep + contr.lin , c.design)
1102 # Wald test
1103 regTermTest(colr6_2, ~gender, df=NULL) # yes
1104
1105 summary(model1)
1106 summary(colr6_2) # model 2
1107 # minimal difference in the beta and SE of both model
1108 # SE in model 2 tends to be slightly higher than model 1
1109
1110
1111 # does adding fast food into model 2 is better?
1112 colr7_1 = svyolr(bmic ~ fastf + gender + softd + edu + age + eth_count + dep + contr.lin , c.design)
1113 regTermTest(colr7_1, ~fastf , df=NULL) # no
1114
1115 # col6_2 is the final model
1116 # one-tailed z-test for deprivation quintiles
1117 pnorm(coefficients(summary(colr6_2))[13,3], lower.tail=FALSE) # quintile 2
1118 pnorm(coefficients(summary(colr6_2))[14,3], lower.tail=FALSE) # quintile 3
1119 pnorm(coefficients(summary(colr6_2))[15,3], lower.tail=FALSE) # quintile 4
1120 pnorm(coefficients(summary(colr6_2))[16,3], lower.tail=FALSE) # quintile 5
1121
1122
1123 ##### Proportional Odds Assumption – Child #####
1124
1125 ## Binomial regression using contrast

```

```

1126 # cex1 is assessing the odds of being in the overweight/obese vs. average category
1127 # cex2 is assessing the odds of being in the obese vs. average/overweight category
1128 cex1 = svyglm(l(bmic>"Average") ~ dep + contr.lin + eth_count + age + edu + softd + gender, c.design
    , family = "binomial")
1129 cex2 = svyglm(l(bmic>"Overweight") ~ dep + contr.lin + eth_count + age + edu + softd + gender, c.
    design, family="binomial")
1130
1131 plot(coef(cex1)[-1], xlim=c(0,20), ylim=c(-1,2), type="l", ylab="Beta", main="Comparison of Betas
    for \nProportional Odds Assumption")
1132 lines(coef(cex2)[-1], col=2)
1133
1134 ## Beta coefficients and standard errors
1135 rbind(coef(cex1),coef(cex2))[, -1]
1136 rbind(SE(cex1),SE(cex2))[, -1]
1137 # higher beta of being in the obese vs average/overweight category compared with being in overweight
    /obese vs average for the following variables: soft drink 4+/week, all ethnicities, and
    deprivation.
1138
1139 ## odds ratios
1140 rbind(exp(coef(cex1)),exp(coef(cex2)))[, -1]
1141
1142 # 2.5% CL for OR
1143 rbind(exp(coef(cex1)-SE(cex1)), exp(coef(cex2)-SE(cex2)))[, -1]
1144
1145 # 97.5% CL for OR
1146 rbind(exp(coef(cex1)+SE(cex1)), exp(coef(cex2)+SE(cex2)))[, -1]

```

Listing A.3: Bivariate Analyses - BMI

```
1
2 ##### BIVARIATE ANALYSES
3 ## This section contains unweighted plots and appropriate statistical tests on all of the
4     combination between the explanatory variables and the outcome variables
5
6 ## Merge
7 ac06 <- merge(a06, c06, all=TRUE)
8 ac11 <- merge(a11, c11, all=TRUE)
9 ac12 <- merge(a12, c12, all=TRUE)
10 ac13 <- merge(a13, c13, all=TRUE)
11 ac14 <- merge(a14, c14, all=TRUE)
12
13 # add year variable to all data sets
14 hs02$year <- rep("2002/03")
15 ac06$year <- rep("2006/07")
16 ac11$year <- rep("2011/12")
17 ac12$year <- rep("2012/13")
18 ac13$year <- rep("2013/14")
19 ac14$year <- rep("2014/15")
20
21 # Split them into child data (2–17 years old) and adult data (18+ years old)
22 adult <- subset(hs02, age >=18) %>%
23     merge(subset(ac06, age >=18), all=TRUE) %>%
24     merge(subset(ac11, age >=18), all=TRUE) %>%
25     merge(subset(ac12, age >=18), all=TRUE) %>%
26     merge(subset(ac13, age >=18), all=TRUE) %>%
27     merge(subset(ac14, age >=18), all=TRUE)
28
29 child <- subset(ac06, age >=2 & age <18) %>%
30     merge(subset(ac11, age >=2 & age <18), all=TRUE) %>%
31     merge(subset(ac12, age >=2 & age <18), all=TRUE) %>%
32     merge(subset(ac13, age >=2 & age <18), all=TRUE) %>%
33     merge(subset(ac14, age >=2 & age <18), all=TRUE)
34
35
36 # Exclude missing BMI values in the Adult data
37 adult <- subset(adult, !is.na(bmiscale))
38
```

```

39
40 # Transform them into tertiles
41 adult$bmi_t <- factor( ntile( adult$bmscale , 3) ,
42                        labels=c( "Bottom Tertile" , "Mid Tertile" , "Top Tertile" ) , ordered=TRUE)
43
44
45 # Include only relevant variables
46 adult <- subset( adult , select = c( "age" , "bmic" , "gender" , "bmscale" , "UA" , "bmi_t" ,
47                                     "dep" , "hhinc" , "edu" , "active" , "native" ,
48                                     "sedentary" , "stair" , "fruit" , "veges" ,
49                                     "year" , "smoke" , "eth_count" , "finalwgt" , "cluster" , "strata" ,
50                                     "haz_drinker_all" ) )
51
52
53 # also exclude missing BMI value in the Child data and include relevant variables
54 child <- subset( child , !is.na( bmic ) )
55
56 child <- subset( child , select=c( "age" , "bmic" , "gender" , "bmscale" , "UA" ,
57                                   "dep" , "hhinc" , "edu" , "fruit" , "veges" ,
58                                   "year" , "softd" , "fastf" , "eth_count" ,
59                                   "finalwgt" , "strata" , "cluster" ) )
60
61
62
63 ## Adult Data
64 ## Subset the data based on year
65 adult02 = subset( adult , year == "2002/03" )
66 adult06 = subset( adult , year == "2006/07" )
67 adult11 = subset( adult , year == "2011/12" )
68 adult12 = subset( adult , year == "2012/13" )
69 adult13 = subset( adult , year == "2013/14" )
70 adult14 = subset( adult , year == "2014/15" )
71
72
73 # AGE
74 grid.arrange( box_plot_median( adult02 , "bmi_t" , "age" , "Age" , "BMI Tertile Group" , "2002/03" ) ,
75               box_plot_median( adult06 , "bmi_t" , "age" , "Age" , "BMI Tertile Group" , "2006/07" ) ,
76               box_plot_median( adult11 , "bmi_t" , "age" , "Age" , "BMI Tertile Group" , "2011/12" ) ,
77               box_plot_median( adult12 , "bmi_t" , "age" , "Age" , "BMI Tertile Group" , "2012/13" ) ,
78               box_plot_median( adult13 , "bmi_t" , "age" , "Age" , "BMI Tertile Group" , "2013/14" ) ,

```

```

79     box_plot_median(adult14 , "bmi_t", "age", "Age", "BMI Tertile Group", "2014/15"),
80     ncol=3, nrow=2, padding = unit(3, "line"))
81
82
83 rbind(anova_table(adult02$age, adult02$bmi_t),
84 anova_table(adult06$age, adult06$bmi_t),
85 anova_table(adult11$age, adult11$bmi_t),
86 anova_table(adult12$age, adult12$bmi_t),
87 anova_table(adult13$age, adult13$bmi_t),
88 anova_table(adult14$age, adult14$bmi_t))
89
90
91 # SEX
92 xt_2x3(adult02$gender, adult02$bmi_t)
93 xt_2x3(adult06$gender, adult06$bmi_t)
94 xt_2x3(adult11$gender, adult11$bmi_t)
95 xt_2x3(adult12$gender, adult12$bmi_t)
96 xt_2x3(adult13$gender, adult13$bmi_t)
97 xt_2x3(adult14$gender, adult14$bmi_t)
98
99 rbind(chix_table(adult02$gender, adult02$bmi_t),
100       chix_table(adult06$gender, adult06$bmi_t),
101       chix_table(adult11$gender, adult11$bmi_t),
102       chix_table(adult12$gender, adult12$bmi_t),
103       chix_table(adult13$gender, adult13$bmi_t),
104       chix_table(adult14$gender, adult14$bmi_t))
105
106 grid.arrange(stacked_plot(adult02, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2002/03", ">=18
107     yo"),
108     stacked_plot(adult06, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2006/07", ">=18
109     yo"),
110     stacked_plot(adult11, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2011/12", ">=18
111     yo"),
112     stacked_plot(adult12, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2012/13", ">=18
113     yo"),
114     stacked_plot(adult13, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2013/14", ">=18
115     yo"),
116     stacked_plot(adult14, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2014/15", ">=18
117     yo"),
118     ncol=3, nrow=2, padding= unit(3, "line"))

```

```

113
114
115 # ETHNICITY
116 table(adult02$eth_count, adult02$bmi_t)
117 table(adult06$eth_count, adult06$bmi_t)
118 table(adult11$eth_count, adult11$bmi_t)
119 table(adult12$eth_count, adult12$bmi_t)
120 table(adult13$eth_count, adult13$bmi_t)
121 table(adult13$eth_count, adult13$bmi_t)
122
123 rbind(chix_table(adult02$eth_count, adult02$bmi_t),
124       chix_table(adult06$eth_count, adult06$bmi_t),
125       chix_table(adult11$eth_count, adult11$bmi_t),
126       chix_table(adult12$eth_count, adult12$bmi_t),
127       chix_table(adult13$eth_count, adult13$bmi_t),
128       chix_table(adult14$eth_count, adult14$bmi_t))
129
130 grid.arrange(stacked_plot(adult02, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2002/03", "
    >=18 yo"),
131              stacked_plot(adult06, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2006/07", "
    >=18 yo"),
132              stacked_plot(adult11, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2011/12", "
    >=18 yo"),
133              stacked_plot(adult12, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2012/13", "
    >=18 yo"),
134              stacked_plot(adult13, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2013/14", "
    >=18 yo"),
135              stacked_plot(adult14, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2014/15", "
    >=18 yo"),
136              ncol=3, nrow=2, padding=unit(3,"line"))
137
138
139 # DEPRIVATION
140 xt_5x3(adult02$Nzdep01_quintiles, adult02$bmi_t)
141 xt_5x3(adult06$nzdep06_quintile, adult06$bmi_t)
142 xt_5x3(adult11$nzdep06_quintile, adult11$bmi_t)
143 xt_5x3(adult12$nzdep06_quintile, adult12$bmi_t)
144 xt_5x3(adult13$nzdep13_quintile, adult13$bmi_t)
145 xt_5x3(adult14$nzdep13_quintile, adult14$bmi_t)
146

```



```

147
148 rbind(chix_table(adult02$Nzdep01_quintiles , adult02$bmi_t),
149       chix_table(adult06$nzdep06_quintile , adult06$bmi_t),
150       chix_table(adult11$nzdep06_quintile , adult11$bmi_t),
151       chix_table(adult12$nzdep06_quintile , adult12$bmi_t),
152       chix_table(adult13$nzdep13_quintile , adult13$bmi_t),
153       chix_table(adult14$nzdep13_quintile , adult14$bmi_t))
154
155 grid.arrange(stacked_plot(adult02 , "Nzdep01_quintiles" , "bmi_t" , "Deprivation Quintile" , "BMI
    Tertile" , "2002/03") ,
156              stacked_plot(adult06 , "nzdep06_quintile" , "bmi_t" , "Deprivation Quintile" , "BMI Tertile"
    , "2006/07") ,
157              stacked_plot(adult11 , "nzdep06_quintile" , "bmi_t" , "Deprivation Quintile" , "BMI Tertile"
    , "2011/12") ,
158              stacked_plot(adult12 , "nzdep06_quintile" , "bmi_t" , "Deprivation Quintile" , "BMI Tertile"
    , "2012/13") ,
159              stacked_plot(adult13 , "nzdep13_quintile" , "bmi_t" , "Deprivation Quintile" , "BMI Tertile"
    , "2013/14") ,
160              stacked_plot(adult14 , "nzdep13_quintile" , "bmi_t" , "Deprivation Quintile" , "BMI Tertile"
    , "2014/15") ,
161              ncol= 3, nrow=2, padding=unit(3,"line"))
162
163
164 # URBAN/RURAL
165 xt_2x3(adult02$UA, adult02$bmi_t)
166 xt_2x3(adult06$UA, adult06$bmi_t)
167 xt_2x3(adult14$UA, adult14$bmi_t)
168
169 rbind(chix_table(adult02$UA, adult02$bmi_t) ,
170       chix_table(adult06$UA, adult06$bmi_t) ,
171       chix_table(adult14$UA, adult14$bmi_t))
172
173 grid.arrange(stacked_plot(adult02 , "UA" , "bmi_t" , "Urban/Rural Area" , "BMI Tertile" , "2002/03") ,
174              stacked_plot(adult06 , "UA" , "bmi_t" , "Urban/Rural Area" , "BMI Tertile" , "2006/07") ,
175              stacked_plot(adult14 , "UA" , "bmi_t" , "Urban/Rural Area" , "BMI Tertile" , "2014/15") ,
176              ncol=3, padding=unit(3,"line"))
177
178
179 # HOUSEHOLD INCOME
180 xt_4x3(adult02$hhinc , adult02$bmi_t)

```

```

181 xt_4x3(adult06$hhinc, adult06$bmi_t)
182 xt_4x3(adult11$hhinc, adult11$bmi_t)
183 xt_4x3(adult12$hhinc, adult12$bmi_t)
184 xt_4x3(adult13$hhinc, adult13$bmi_t)
185 xt_4x3(adult14$hhinc, adult14$bmi_t)
186
187 rbind(chix_table(adult02$hhinc, adult02$bmi_t),
188       chix_table(adult06$hhinc, adult06$bmi_t),
189       chix_table(adult11$hhinc, adult11$bmi_t),
190       chix_table(adult12$hhinc, adult12$bmi_t),
191       chix_table(adult13$hhinc, adult13$bmi_t),
192       chix_table(adult14$hhinc, adult14$bmi_t))
193
194 grid.arrange(stacked_plot(adult02, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2002/03
    "),
195              stacked_plot(adult06, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2006/07"
196                           ),
197              stacked_plot(adult11, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2011/12"
198                           ),
199              stacked_plot(adult12, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2012/13"
200                           ),
201              stacked_plot(adult13, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2013/14"
202                           ),
203              stacked_plot(adult14, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2014/15"
204                           ),
205              ncol= 3, nrow=2, padding=unit(3,"line"))
206
207
208
209
210
211
212 # EDUCATION
213 xt_3x3(adult02$edu, adult02$bmi_t)
214 xt_3x3(adult06$edu, adult06$bmi_t)
215 xt_3x3(adult11$edu, adult11$bmi_t)
216 xt_3x3(adult12$edu, adult12$bmi_t)
217 xt_3x3(adult13$edu, adult13$bmi_t)
218 xt_3x3(adult14$edu, adult14$bmi_t)
219
220 rbind(chix_table(adult02$edu, adult02$bmi_t),
221       chix_table(adult06$edu, adult06$bmi_t),
222       chix_table(adult11$edu, adult11$bmi_t),

```

```

215     chix_table(adult12$edu, adult12$bmi_t),
216     chix_table(adult13$edu, adult13$bmi_t),
217     chix_table(adult14$edu, adult14$bmi_t))
218
219 grid.arrange(stacked_plot(adult02, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2002
    /03"),
220               stacked_plot(adult06, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2006/
    07"),
221               stacked_plot(adult11, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2011/
    12"),
222               stacked_plot(adult12, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2012/
    13"),
223               stacked_plot(adult13, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2013/
    14"),
224               stacked_plot(adult14, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2014/
    15"),
225               ncol= 3, nrow=2, padding=unit(3,"line"))
226
227
228 # FRUIT GUIDELINE
229 xt_2x3(adult02$fruit, adult02$bmi_t)
230 xt_2x3(adult06$fruit, adult06$bmi_t)
231 xt_2x3(adult11$fruit, adult11$bmi_t)
232 xt_2x3(adult12$fruit, adult12$bmi_t)
233 xt_2x3(adult13$fruit, adult13$bmi_t)
234 xt_2x3(adult14$fruit, adult14$bmi_t)
235
236 rbind(chix_table(adult02$fruit, adult02$bmi_t),
237       chix_table(adult06$fruit, adult06$bmi_t),
238       chix_table(adult11$fruit, adult11$bmi_t),
239       chix_table(adult12$fruit, adult12$bmi_t),
240       chix_table(adult13$fruit, adult13$bmi_t),
241       chix_table(adult14$fruit, adult14$bmi_t))
242
243 grid.arrange(stacked_plot(adult02, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2002/03"),
244               stacked_plot(adult06, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2006/07"),
245               stacked_plot(adult11, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2011/12"),
246               stacked_plot(adult12, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2012/13"),
247               stacked_plot(adult13, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2013/14"),
248               stacked_plot(adult14, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2014/15"),

```

```

249         ncol= 3, nrow=2, padding=unit(3,"line"))
250
251
252 # VEGETABLE GUIDELINE
253 xt_2x3(adult02$veges, adult02$bmi_t)
254 xt_2x3(adult06$veges, adult06$bmi_t)
255 xt_2x3(adult11$veges, adult11$bmi_t)
256 xt_2x3(adult12$veges, adult12$bmi_t)
257 xt_2x3(adult13$veges, adult13$bmi_t)
258 xt_2x3(adult14$veges, adult14$bmi_t)
259
260 rbind(chix_table(adult02$veges, adult02$bmi_t),
261       chix_table(adult06$veges, adult06$bmi_t),
262       chix_table(adult11$veges, adult11$bmi_t),
263       chix_table(adult12$veges, adult12$bmi_t),
264       chix_table(adult13$veges, adult13$bmi_t),
265       chix_table(adult14$veges, adult14$bmi_t))
266
267 grid.arrange(stacked_plot(adult02, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2002/03"
268   ),
269   stacked_plot(adult06, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2006/07")
270   ,
271   stacked_plot(adult11, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2011/12")
272   ,
273   stacked_plot(adult12, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2012/13")
274   ,
275   stacked_plot(adult13, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2013/14")
276   ,
277   stacked_plot(adult14, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2014/15")
278   ,
279   ncol= 3, nrow=2, padding=unit(3,"line"))
280
281 # MIGRATION
282 xt_2x3(adult02$native, adult02$bmi_t)
283 xt_2x3(adult06$native, adult06$bmi_t)
284 xt_2x3(adult11$native, adult11$bmi_t)
285 xt_2x3(adult12$native, adult12$bmi_t)
286 test <- xt_2x3(adult13$native, adult13$bmi_t)
287 xt_2x3(adult14$native, adult14$bmi_t)

```

```

283
284 rbind(chix_table(adult02$native, adult02$bmi_t),
285        chix_table(adult06$native, adult06$bmi_t),
286        chix_table(adult11$native, adult11$bmi_t),
287        chix_table(adult12$native, adult12$bmi_t),
288        chix_table(adult13$native, adult13$bmi_t),
289        chix_table(adult14$native, adult14$bmi_t))
290
291 grid.arrange(stacked_plot(adult02, "native", "bmi_t", "Migration Status", "BMI Tertile", "2002/03"),
292              stacked_plot(adult06, "native", "bmi_t", "Migration Status", "BMI Tertile", "2006/07"),
293              stacked_plot(adult11, "native", "bmi_t", "Migration Status", "BMI Tertile", "2011/12"),
294              stacked_plot(adult12, "native", "bmi_t", "Migration Status", "BMI Tertile", "2012/13"),
295              stacked_plot(adult13, "native", "bmi_t", "Migration Status", "BMI Tertile", "2013/14"),
296              stacked_plot(adult14, "native", "bmi_t", "Migration Status", "BMI Tertile", "2014/15"),
297              ncol= 3, nrow=2, padding=unit(3,"line"))
298
299
300 # DIFFICULTY CLIMBING STAIRS
301 xt_3x3(adult02$stair, adult02$bmi_t)
302 xt_3x3(adult06$stair, adult06$bmi_t)
303 xt_3x3(adult11$stair, adult11$bmi_t)
304 xt_3x3(adult12$stair, adult12$bmi_t)
305 xt_3x3(adult13$stair, adult13$bmi_t)
306 xt_3x3(adult14$stair, adult14$bmi_t)
307
308 rbind(chix_table(adult02$stair, adult02$bmi_t),
309        chix_table(adult06$stair, adult06$bmi_t),
310        chix_table(adult11$stair, adult11$bmi_t),
311        chix_table(adult12$stair, adult12$bmi_t),
312        chix_table(adult13$stair, adult13$bmi_t),
313        chix_table(adult14$stair, adult14$bmi_t))
314
315 grid.arrange(stacked_plot(adult02, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
    2002/03"),
316              stacked_plot(adult06, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
    2006/07"),
317              stacked_plot(adult11, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
    2011/12"),
318              stacked_plot(adult12, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
    2012/13"),

```

```

319         stacked_plot(adult13, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
320             2013/14"),
321         stacked_plot(adult14, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
322             2014/15"),
323
324         ncol= 3, nrow=2, padding=unit(3,"line"))
325
326 # SMOKING STATUS
327
328 xt_3x3(adult06$smoke, adult06$bmi_t)
329 xt_3x3(adult11$smoke, adult11$bmi_t)
330 xt_3x3(adult12$smoke, adult12$bmi_t)
331 xt_3x3(adult13$smoke, adult13$bmi_t)
332 xt_3x3(adult14$smoke, adult14$bmi_t)
333
334 rbind(chix_table(adult06$smoke, adult06$bmi_t),
335        chix_table(adult11$smoke, adult11$bmi_t),
336        chix_table(adult12$smoke, adult12$bmi_t),
337        chix_table(adult13$smoke, adult13$bmi_t),
338        chix_table(adult14$smoke, adult14$bmi_t))
339
340 grid.arrange(stacked_plot(adult06, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2006/07"),
341              stacked_plot(adult11, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2011/12"),
342              stacked_plot(adult12, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2012/13"),
343              stacked_plot(adult13, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2013/14"),
344              stacked_plot(adult14, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2014/15"),
345              ncol= 3, nrow=2, padding=unit(3,"line"))
346
347 # ALCOHOL PROBLEM
348
349 xt_2x3(adult02$haz_drinker_all, adult02$bmi_t)
350 xt_2x3(adult06$haz_drinker_all, adult06$bmi_t)
351 xt_2x3(adult11$haz_drinker_all, adult11$bmi_t)
352 xt_2x3(adult12$haz_drinker_all, adult12$bmi_t)
353 xt_2x3(adult13$haz_drinker_all, adult13$bmi_t)
354 xt_2x3(adult14$haz_drinker_all, adult14$bmi_t)
355
356 rbind(chix_table(adult02$haz_drinker_all, adult02$bmi_t),
357        chix_table(adult06$haz_drinker_all, adult06$bmi_t),
358        chix_table(adult11$haz_drinker_all, adult11$bmi_t),
359        chix_table(adult12$haz_drinker_all, adult12$bmi_t),

```

```

357     chix_table(adult13$haz_drinker_all, adult13$bmi_t),
358     chix_table(adult14$haz_drinker_all, adult14$bmi_t))
359
360 grid.arrange(stacked_plot(adult02, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
    2002/03"),
361               stacked_plot(adult06, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
    2006/07"),
362               stacked_plot(adult11, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
    2011/12"),
363               stacked_plot(adult12, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
    2012/13"),
364               stacked_plot(adult13, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
    2013/14"),
365               stacked_plot(adult14, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
    2014/15"),
366               ncol= 3, nrow=2, padding=unit(3,"line"))
367
368
369 # PHYSICAL ACTIVITY
370 xt_2x3(adult02$active, adult02$bmi_t)
371 xt_2x3(adult06$active, adult06$bmi_t)
372 xt_2x3(adult11$active, adult11$bmi_t)
373 xt_2x3(adult12$active, adult12$bmi_t)
374 xt_2x3(adult13$active, adult13$bmi_t)
375 xt_2x3(adult14$active, adult14$bmi_t)
376
377 rbind(chix_table(adult02$active, adult02$bmi_t),
378       chix_table(adult06$active, adult06$bmi_t),
379       chix_table(adult11$active, adult11$bmi_t),
380       chix_table(adult12$active, adult12$bmi_t),
381       chix_table(adult13$active, adult13$bmi_t),
382       chix_table(adult14$active, adult14$bmi_t))
383
384 grid.arrange(stacked_plot(adult02, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2002/03")
    ,
385               stacked_plot(adult06, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2006/07"),
386               stacked_plot(adult11, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2011/12"),
387               stacked_plot(adult12, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2012/13"),
388               stacked_plot(adult13, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2013/14"),
389               stacked_plot(adult14, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2014/15"),

```

```

390         ncol= 3, nrow=2, padding=unit(3,"line"))
391
392
393 # SEDENTARY LIFESTYLE
394 xt_2x3(adult02$sedentary, adult02$bmi_t)
395 xt_2x3(adult06$sedentary, adult06$bmi_t)
396 xt_2x3(adult11$sedentary, adult11$bmi_t)
397 xt_2x3(adult12$sedentary, adult12$bmi_t)
398 xt_2x3(adult13$sedentary, adult13$bmi_t)
399 xt_2x3(adult14$sedentary, adult14$bmi_t)
400
401 rbind(chix_table(adult02$sedentary, adult02$bmi_t),
402       chix_table(adult06$sedentary, adult06$bmi_t),
403       chix_table(adult11$sedentary, adult11$bmi_t),
404       chix_table(adult12$sedentary, adult12$bmi_t),
405       chix_table(adult13$sedentary, adult13$bmi_t),
406       chix_table(adult14$sedentary, adult14$bmi_t))
407
408 grid.arrange(stacked_plot(adult02, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2002
409 /03"),
410              stacked_plot(adult06, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2006/
411 07"),
412              stacked_plot(adult11, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2011/
413 12"),
414              stacked_plot(adult12, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2012/
415 13"),
416              stacked_plot(adult13, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2013/
417 14"),
418              stacked_plot(adult14, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2014/
419 15"),
420              ncol= 3, nrow=2, padding=unit(3,"line"))
421
422 ## CHILD DATA
423 ## Subset the child data on each year
424 child06 = subset(child, year == "2006/07")
425 child11 = subset(child, year == "2011/12")
426 child12 = subset(child, year == "2012/13")
427 child13 = subset(child, year == "2013/14")
428 child14 = subset(child, year == "2014/15")

```



```

424
425
426 # AGE
427 grid.arrange(box_plot_median(child06, "bmic", "age", "Age", "BMI Category", "2006/07", max= 20, bins
    =5, txt.size = 4),
428             box_plot_median(child11, "bmic", "age", "Age", "BMI Category", "2011/12", max= 20, bins
    =5, txt.size = 4),
429             box_plot_median(child12, "bmic", "age", "Age", "BMI Category", "2012/13", max= 20, bins
    =5, txt.size = 4),
430             box_plot_median(child13, "bmic", "age", "Age", "BMI Category", "2013/14", max= 20, bins
    =5, txt.size = 4),
431             box_plot_median(child14, "bmic", "age", "Age", "BMI Category", "2014/15", max= 20, bins
    =5, txt.size = 4),
432             ncol=3, nrow=2, padding = unit(3, "line"))
433
434
435 rbind(anova_table(child06$age, child06$bmic),
436 anova_table(child11$age, child11$bmic),
437 anova_table(child12$age, child12$bmic),
438 anova_table(child13$age, child13$bmic),
439 anova_table(child14$age, child14$bmic))
440
441
442
443 # SEX
444 xt_2x3(child06$gender, child06$bmic)
445 xt_2x3(child11$gender, child11$bmic)
446 xt_2x3(child12$gender, child12$bmic)
447 xt_2x3(child13$gender, child13$bmic)
448 xt_2x3(child14$gender, child14$bmic)
449
450 rbind(chix_table(child06$gender, child06$bmic),
451       chix_table(child11$gender, child11$bmic),
452       chix_table(child12$gender, child12$bmic),
453       chix_table(child13$gender, child13$bmic),
454       chix_table(child14$gender, child14$bmic))
455
456 grid.arrange(stacked_plot(child06, "gender", "bmic", "Sex Group", "BMI Category", "2006/07", "2—<18
    yo"),
457             stacked_plot(child11, "gender", "bmic", "Sex Group", "BMI Category", "2011/12", "2—<18

```

```

    yo"),
458 stacked_plot(child12, "gender", "bmic", "Sex Group", "BMI Category", "2012/13", "2--<18
    yo"),
459 stacked_plot(child13, "gender", "bmic", "Sex Group", "BMI Category", "2013/14", "2--<18
    yo"),
460 stacked_plot(child14, "gender", "bmic", "Sex Group", "BMI Category", "2014/15", "2--<18
    yo"),
461 ncol=3, nrow=2, padding= unit(3, "line"))
462
463
464 # ETHNICITY
465 table(child06$eth_count, child06$bmic)
466 table(child11$eth_count, child11$bmic)
467 table(child12$eth_count, child12$bmic)
468 table(child13$eth_count, child13$bmic)
469 table(child14$eth_count, child14$bmic)
470
471 rbind(chix_table(child06$eth_count, child06$bmic),
472       chix_table(child11$eth_count, child11$bmic),
473       chix_table(child12$eth_count, child12$bmic),
474       chix_table(child13$eth_count, child13$bmic),
475       chix_table(child14$eth_count, child14$bmic))
476
477 grid.arrange(stacked_plot(child06, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2006/07",
    "2--<18 yo"),
478             stacked_plot(child11, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2011/12",
    "2--<18 yo"),
479             stacked_plot(child12, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2012/13",
    "2--<18 yo"),
480             stacked_plot(child13, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2013/14",
    "2--<18 yo"),
481             stacked_plot(child14, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2014/15",
    "2--<18 yo"),
482             ncol=3, nrow=2, padding= unit(3, "line"))
483
484
485 # DEPRIVATION
486 xt_5x3(child06$dep, child06$bmic)
487 xt_5x3(child11$dep, child11$bmic)
488 xt_5x3(child12$dep, child12$bmic)

```

```

489 xt_5x3(child13$dep, child13$bmic)
490 xt_5x3(child14$dep, child14$bmic)
491
492 rbind(chix_table(child06$dep, child06$bmic),
493       chix_table(child11$dep, child11$bmic),
494       chix_table(child12$dep, child12$bmic),
495       chix_table(child13$dep, child13$bmic),
496       chix_table(child14$dep, child14$bmic))
497
498 grid.arrange(stacked_plot(child06, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2006/07"
    , "2--<18 yo"),
499             stacked_plot(child11, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2011/12"
    , "2--<18 yo"),
500             stacked_plot(child12, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2012/13"
    , "2--<18 yo"),
501             stacked_plot(child13, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2013/14"
    , "2--<18 yo"),
502             stacked_plot(child14, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2014/15"
    , "2--<18 yo"),
503             ncol=3, nrow=2, padding= unit(3, "line"))
504
505
506 # URBAN/RURAL AREA
507 xt_2x3(child06$UA, child06$bmic)
508 xt_2x3(child14$UA, child14$bmic)
509
510 rbind(chix_table(child06$UA, child06$bmic),
511       chix_table(child14$UA, child14$bmic))
512
513 grid.arrange(stacked_plot(child06, "UA", "bmic", "Urban/Rural Area", "BMI Category", "2006/07", "
    2--<18 yo"),
514             stacked_plot(child14, "UA", "bmic", "Urban/Rural Area", "BMI Category", "2014/15", "
    2--<18 yo"),
515             ncol=3, nrow=2, padding= unit(3, "line"))
516
517
518 # HOUSEHOLD INCOME
519 xt_4x3(child06$hhinc, child06$bmic)
520 xt_4x3(child11$hhinc, child11$bmic)
521 xt_4x3(child12$hhinc, child12$bmic)

```

```

522 xt_4x3(child13$hhinc, child13$bmic)
523 xt_4x3(child14$hhinc, child14$bmic)
524
525 rbind(chix_table(child06$hhinc, child06$bmic),
526       chix_table(child11$hhinc, child11$bmic),
527       chix_table(child12$hhinc, child12$bmic),
528       chix_table(child13$hhinc, child13$bmic),
529       chix_table(child14$hhinc, child14$bmic))
530
531 grid.arrange(stacked_plot(child06, "hhinc", "bmic", "Household Income", "BMI Category", "2006/07", "
    2--<18 yo"),
532              stacked_plot(child11, "hhinc", "bmic", "Household Income", "BMI Category", "2011/12", "
    2--<18 yo"),
533              stacked_plot(child12, "hhinc", "bmic", "Household Income", "BMI Category", "2012/13", "
    2--<18 yo"),
534              stacked_plot(child13, "hhinc", "bmic", "Household Income", "BMI Category", "2013/14", "
    2--<18 yo"),
535              stacked_plot(child14, "hhinc", "bmic", "Household Income", "BMI Category", "2014/15", "
    2--<18 yo"),
536              ncol=3, nrow=2, padding= unit(3, "line"))
537
538
539
540 # EDUCATION
541 xt_3x3(child06$edu, child06$bmic)
542 xt_3x3(child11$edu, child11$bmic)
543 xt_3x3(child12$edu, child12$bmic)
544 xt_3x3(child13$edu, child13$bmic)
545 xt_3x3(child14$edu, child14$bmic)
546
547 rbind(chix_table(child06$edu, child06$bmic),
548       chix_table(child11$edu, child11$bmic),
549       chix_table(child12$edu, child12$bmic),
550       chix_table(child13$edu, child13$bmic),
551       chix_table(child14$edu, child14$bmic))
552
553 grid.arrange(stacked_plot(child06, "edu", "bmic", "Educational Qualification", "BMI Category", "2006
    /07", "2--<18 yo"),
554              stacked_plot(child11, "edu", "bmic", "Educational Qualification", "BMI Category", "2011
    /12", "2--<18 yo"),

```

```

555 stacked_plot(child12, "edu", "bmic", "Educational Qualification", "BMI Category", "2012
    /13", "2--<18 yo"),
556 stacked_plot(child13, "edu", "bmic", "Educational Qualification", "BMI Category", "2013
    /14", "2--<18 yo"),
557 stacked_plot(child14, "edu", "bmic", "Educational Qualification", "BMI Category", "2014
    /15", "2--<18 yo"),
558 ncol=3, nrow=2, padding= unit(3, "line"))
559
560
561
562 # FRUIT GUIDELINE
563 xt_2x3(child06$fruit, child06$bmic)
564 xt_2x3(child11$fruit, child11$bmic)
565 xt_2x3(child12$fruit, child12$bmic)
566 xt_2x3(child13$fruit, child13$bmic)
567 xt_2x3(child14$fruit, child14$bmic)
568
569 rbind(chix_table(child06$fruit, child06$bmic),
570       chix_table(child11$fruit, child11$bmic),
571       chix_table(child12$fruit, child12$bmic),
572       chix_table(child13$fruit, child13$bmic),
573       chix_table(child14$fruit, child14$bmic))
574
575 grid.arrange(stacked_plot(child06, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2006/07", "
    2--<18 yo"),
576              stacked_plot(child11, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2011/12", "
    2--<18 yo"),
577              stacked_plot(child12, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2012/13", "
    2--<18 yo"),
578              stacked_plot(child13, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2013/14", "
    2--<18 yo"),
579              stacked_plot(child14, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2014/15", "
    2--<18 yo"),
580              ncol=3, nrow=2, padding= unit(3, "line"))
581
582
583
584 # VEGETABLE GUIDELINE
585 xt_2x3(child06$veges, child06$bmic)
586 xt_2x3(child11$veges, child11$bmic)

```

```

587 xt_2x3(child12$veges, child12$bmic)
588 xt_2x3(child13$veges, child13$bmic)
589 xt_2x3(child14$veges, child14$bmic)
590
591 rbind(chix_table(child06$veges, child06$bmic),
592       chix_table(child11$veges, child11$bmic),
593       chix_table(child12$veges, child12$bmic),
594       chix_table(child13$veges, child13$bmic),
595       chix_table(child14$veges, child14$bmic))
596
597 grid.arrange(stacked_plot(child06, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2006/07"
598   , "2--<18 yo"),
599   stacked_plot(child11, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2011/12"
600   , "2--<18 yo"),
601   stacked_plot(child12, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2012/13"
602   , "2--<18 yo"),
603   stacked_plot(child13, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2013/14"
604   , "2--<18 yo"),
605   stacked_plot(child14, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2014/15"
606   , "2--<18 yo"),
607   ncol=3, nrow=2, padding= unit(3, "line"))
608
609 # SOFT DRINK CONSUMPTION
610
611 xt_4x3(child06$softd, child06$bmic)
612 xt_4x3(child11$softd, child11$bmic)
613 xt_4x3(child12$softd, child12$bmic)
614 xt_4x3(child13$softd, child13$bmic)
615 xt_4x3(child14$softd, child14$bmic)
616
617 rbind(chix_table(child06$softd, child06$bmic),
618       chix_table(child11$softd, child11$bmic),
619       chix_table(child12$softd, child12$bmic),
620       chix_table(child13$softd, child13$bmic),
621       chix_table(child14$softd, child14$bmic))
622
623 grid.arrange(stacked_plot(child06, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2006/
624   07", "2--<18 yo"),
625   stacked_plot(child11, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2011/

```

```

12", "2--<18 yo"),
621 stacked_plot(child12, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2012/
13", "2--<18 yo"),
622 stacked_plot(child13, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2013/
14", "2--<18 yo"),
623 stacked_plot(child14, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2014/
15", "2--<18 yo"),
624 ncol=3, nrow=2, padding= unit(3, "line"))
625
626
627 # FAST FOOD CONSUMPTION
628 xt_4x3(child06$fastf, child06$bmic)
629 xt_4x3(child11$fastf, child11$bmic)
630 xt_4x3(child12$fastf, child12$bmic)
631 xt_4x3(child13$fastf, child13$bmic)
632 xt_4x3(child14$fastf, child14$bmic)
633
634 rbind(chix_table(child06$fastf, child06$bmic),
635        chix_table(child11$fastf, child11$bmic),
636        chix_table(child12$fastf, child12$bmic),
637        chix_table(child13$fastf, child13$bmic),
638        chix_table(child14$fastf, child14$bmic))
639
640 grid.arrange(stacked_plot(child06, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2006/
07", "2--<18 yo"),
641              stacked_plot(child11, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2011/
12", "2--<18 yo"),
642              stacked_plot(child12, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2012/
13", "2--<18 yo"),
643              stacked_plot(child13, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2013/
14", "2--<18 yo"),
644              stacked_plot(child14, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2014/
15", "2--<18 yo"),
645              ncol=3, nrow=2, padding= unit(3, "line"))

```

Listing A.4: Bivariate Analyses - Deprivation Quintiles

```
1
2 ##### BIVARIATE ANALYSES
3 ## Examining confounding factors on the deprivation quintile variable
4
5
6 ## Adult Data
7 # Age
8 rbind(anova_table(adult02$age, adult02$dep),
9       anova_table(adult06$age, adult06$dep),
10      anova_table(adult11$age, adult11$dep),
11      anova_table(adult12$age, adult12$dep),
12      anova_table(adult13$age, adult13$dep),
13      anova_table(adult14$age, adult14$dep))
14
15
16 # Sex
17 rbind(chix_table(adult02$gender, adult02$dep),
18       chix_table(adult06$gender, adult06$dep),
19       chix_table(adult11$gender, adult11$dep),
20       chix_table(adult12$gender, adult12$dep),
21       chix_table(adult13$gender, adult13$dep),
22       chix_table(adult14$gender, adult14$dep))
23
24
25 # Urban/Rural area
26 rbind(chix_table(adult02$UA, adult02$dep),
27       chix_table(adult06$UA, adult06$dep),
28       chix_table(adult14$UA, adult14$dep))
29
30
31 # Household Income
32 rbind(chix_table(adult02$hhinc, adult02$dep),
33       chix_table(adult06$hhinc, adult06$dep),
34       chix_table(adult11$hhinc, adult11$dep),
35       chix_table(adult12$hhinc, adult12$dep),
36       chix_table(adult13$hhinc, adult13$dep),
37       chix_table(adult14$hhinc, adult14$dep))
38
39
```



```

40 # Education
41 rbind(chix_table(adult02$edu, adult02$dep),
42       chix_table(adult06$edu, adult06$dep),
43       chix_table(adult11$edu, adult11$dep),
44       chix_table(adult12$edu, adult12$dep),
45       chix_table(adult13$edu, adult13$dep),
46       chix_table(adult14$edu, adult14$dep))
47
48
49 # Fruit Guideline
50 rbind(chix_table(adult02$fruit, adult02$dep),
51       chix_table(adult06$fruit, adult06$dep),
52       chix_table(adult11$fruit, adult11$dep),
53       chix_table(adult12$fruit, adult12$dep),
54       chix_table(adult13$fruit, adult13$dep),
55       chix_table(adult14$fruit, adult14$dep))
56
57
58 # Vegetable Guideline
59 rbind(chix_table(adult02$veges, adult02$dep),
60       chix_table(adult06$veges, adult06$dep),
61       chix_table(adult11$veges, adult11$dep),
62       chix_table(adult12$veges, adult12$dep),
63       chix_table(adult13$veges, adult13$dep),
64       chix_table(adult14$veges, adult14$dep))
65
66
67 # Migration Status
68 rbind(chix_table(adult02$native, adult02$dep),
69       chix_table(adult06$native, adult06$dep),
70       chix_table(adult11$native, adult11$dep),
71       chix_table(adult12$native, adult12$dep),
72       chix_table(adult13$native, adult13$dep),
73       chix_table(adult14$native, adult14$dep))
74
75
76 # Difficulty Climbing Stairs
77 rbind(chix_table(adult02$stair, adult02$dep),
78       chix_table(adult06$stair, adult06$dep),
79       chix_table(adult11$stair, adult11$dep),

```

```

80     chix_table(adult12$stair, adult12$dep),
81     chix_table(adult13$stair, adult13$dep),
82     chix_table(adult14$stair, adult14$dep))
83
84
85 # Smoking Status
86 rbind(chix_table(adult02$smoke, adult02$dep),
87       chix_table(adult06$smoke, adult06$dep),
88       chix_table(adult11$smoke, adult11$dep),
89       chix_table(adult12$smoke, adult12$dep),
90       chix_table(adult13$smoke, adult13$dep),
91       chix_table(adult14$smoke, adult14$dep))
92
93
94 # Alcohol Problem
95 rbind(chix_table(adult02$haz_drinker_all, adult02$dep),
96       chix_table(adult06$haz_drinker_all, adult06$dep),
97       chix_table(adult11$haz_drinker_all, adult11$dep),
98       chix_table(adult12$haz_drinker_all, adult12$dep),
99       chix_table(adult13$haz_drinker_all, adult13$dep),
100      chix_table(adult14$haz_drinker_all, adult14$dep))
101
102
103 # Physical Acitivity
104 rbind(chix_table(adult02$active, adult02$dep),
105       chix_table(adult06$active, adult06$dep),
106       chix_table(adult11$active, adult11$dep),
107       chix_table(adult12$active, adult12$dep),
108       chix_table(adult13$active, adult13$dep),
109       chix_table(adult14$active, adult14$dep))
110
111
112 # Sedentary Lifestyle
113 rbind(chix_table(adult02$sedentary, adult02$dep),
114       chix_table(adult06$sedentary, adult06$dep),
115       chix_table(adult11$sedentary, adult11$dep),
116       chix_table(adult12$sedentary, adult12$dep),
117       chix_table(adult13$sedentary, adult13$dep),
118       chix_table(adult14$sedentary, adult14$dep))
119

```

```

120
121
122 ## Child Data
123 # Age
124 rbind(anova_table(child06$age, child06$dep),
125       anova_table(child11$age, child11$dep),
126       anova_table(child12$age, child12$dep),
127       anova_table(child13$age, child13$dep),
128       anova_table(child14$age, child14$dep))
129
130
131 # Sex
132 rbind(chix_table(child06$gender, child06$dep),
133       chix_table(child11$gender, child11$dep),
134       chix_table(child12$gender, child12$dep),
135       chix_table(child13$gender, child13$dep),
136       chix_table(child14$gender, child14$dep))
137
138
139 # Urban/Rural area
140 rbind(chix_table(child06$UA, child06$dep),
141       chix_table(child14$UA, child14$dep))
142
143
144 # Household Income
145 rbind(chix_table(child06$hhinc, child06$dep),
146       chix_table(child11$hhinc, child11$dep),
147       chix_table(child12$hhinc, child12$dep),
148       chix_table(child13$hhinc, child13$dep),
149       chix_table(child14$hhinc, child14$dep))
150
151
152 # Education
153 rbind(chix_table(child06$edu, child06$dep),
154       chix_table(child11$edu, child11$dep),
155       chix_table(child12$edu, child12$dep),
156       chix_table(child13$edu, child13$dep),
157       chix_table(child14$edu, child14$dep))
158
159

```

```

160 # softd Guideline
161 rbind(chix_table(child06$fruit, child06$dep),
162       chix_table(child11$fruit, child11$dep),
163       chix_table(child12$fruit, child12$dep),
164       chix_table(child13$fruit, child13$dep),
165       chix_table(child14$fruit, child14$dep))
166
167
168 # Vegetable Guideline
169 rbind(chix_table(child06$veges, child06$dep),
170       chix_table(child11$veges, child11$dep),
171       chix_table(child12$veges, child12$dep),
172       chix_table(child13$veges, child13$dep),
173       chix_table(child14$veges, child14$dep))
174
175
176 # Soft Drink Consumption
177 rbind(chix_table(child06$softd, child06$dep),
178       chix_table(child11$softd, child11$dep),
179       chix_table(child12$softd, child12$dep),
180       chix_table(child13$softd, child13$dep),
181       chix_table(child14$softd, child14$dep))
182
183
184 # Fast Food Consumption
185 rbind(chix_table(child06$fastf, child06$dep),
186       chix_table(child11$fastf, child11$dep),
187       chix_table(child12$fastf, child12$dep),
188       chix_table(child13$fastf, child13$dep),
189       chix_table(child14$fastf, child14$dep))

```

B Bivariate Tables

B.1 Bivariate Tables for The Adult Data

Table B.1: Mean Age by BMI Tertiles

BMI Tertiles	Age (in years)
	mean (SD)
Bottom Tertile	46.08 (18.92)
Mid Tertile	50.36 (17.83)
Top Tertile	49.29 (16.45)

Table B.2: Sex by BMI Tertiles

BMI Tertiles	Male	Female
	N (row %)	
Bottom Tertile	8570 (0.38)	14007 (0.62)
Mid Tertile	11539 (0.51)	11038 (0.49)
Top Tertile	9237 (0.41)	13340 (0.59)

Table B.3: Ethnicities by BMI Tertiles

BMI Tertiles	European Only	2+ Ethnicities (M)	Asian Only	Maori Only	Pacific Only	Other
	N (row %)					
Bottom Tertile	13645 (0.61)	1619 (0.07)	1997 (0.09)	1783 (0.08)	246 (0.01)	3230 (0.14)
Mid Tertile	13719 (0.61)	1855 (0.08)	1058 (0.05)	2714 (0.12)	715 (0.03)	2452 (0.11)
Top Tertile	10534 (0.47)	2337 (0.10)	373 (0.02)	5167 (0.23)	2320 (0.10)	1804 (0.08)

Table B.4: Deprivation Quintiles by BMI Tertiles

BMI Tertiles	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
	N (row %)				
Bottom Tertile	3779 (0.17)	4058 (0.18)	4688 (0.21)	5043 (0.22)	5006 (0.22)
Mid Tertile	3507 (0.16)	3950 (0.18)	4414 (0.20)	5053 (0.22)	5640 (0.25)
Top Tertile	2233 (0.10)	2693 (0.12)	3781 (0.17)	5181 (0.23)	8678 (0.38)

Table B.5: Urban/Rural Area by BMI Tertiles

BMI Tertiles	Urban	Rural
	N (row %)	
Bottom Tertile	9563 (0.83)	1910 (0.17)
Mid Tertile	9124 (0.81)	2175 (0.19)
Top Tertile	8868 (0.79)	2367 (0.21)

Table B.6: Household Income by BMI Tertiles

BMI Tertiles	≤\$15,000	\$15,001-\$40,000	\$40,001-\$70,000	>\$70,000
	N (row %)			
Bottom Tertile	1452 (0.08)	5280 (0.31)	4325 (0.25)	6106 (0.36)
Mid Tertile	1362 (0.08)	5478 (0.31)	4475 (0.25)	6639 (0.37)
Top Tertile	1716 (0.10)	5942 (0.34)	4273 (0.25)	5442 (0.31)

Table B.7: Educational Qualifications by BMI Tertiles

BMI Tertiles	No Qualification	Secondary	Tertiary
	N (row %)		
Bottom Tertile	3623 (0.16)	6866 (0.31)	11923 (0.53)
Mid Tertile	4576 (0.20)	6487 (0.29)	11363 (0.51)
Top Tertile	5988 (0.27)	7044 (0.32)	9319 (0.42)

Table B.8: Adherence to Fruit Guideline by BMI Tertiles

BMI Tertiles	No	Yes
	N (row %)	
Bottom Tertile	9610 (0.43)	12956 (0.57)
Mid Tertile	9762 (0.43)	12807 (0.57)
Top Tertile	10298 (0.46)	12267 (0.54)

Table B.9: Adherence to Vegetable Guideline by BMI Tertiles

BMI Tertiles	No	Yes
	N (row %)	
Bottom Tertile	8892 (0.39)	13674 (0.61)
Mid Tertile	8609 (0.38)	13956 (0.62)
Top Tertile	9139 (0.41)	13422 (0.59)

Table B.10: Migration Status by BMI Tertiles

BMI Tertiles	Migrant	Native
	N (row %)	
Bottom Tertile	2576 (0.11)	19990 (0.89)
Mid Tertile	1489 (0.07)	21077 (0.93)
Top Tertile	985 (0.04)	21574 (0.96)

Table B.11: Difficulty Climbing Stairs by BMI Tertiles

BMI Tertiles	No Difficulty	A Little Difficult	A Lot Difficult
	N (row %)		
Bottom Tertile	18708 (0.83)	2297 (0.10)	1522 (0.07)
Mid Tertile	17718 (0.79)	2930 (0.13)	1888 (0.08)
Top Tertile	15251 (0.68)	4165 (0.18)	3121 (0.14)

Table B.12: Smoking Status by BMI Tertiles

BMI Tertiles	Non Smoker	Ex Smoker	Current Smoker
	N (row %)		
Bottom Tertile	10296 (0.55)	4201 (0.22)	4260 (0.23)
Mid Tertile	9395 (0.50)	5626 (0.30)	3949 (0.21)
Top Tertile	8483 (0.44)	6127 (0.32)	4475 (0.23)

Table B.13: Drinking Problem by BMI Tertiles

BMI Tertiles	No Drinking Problem	Drinking Problem
	N (row %)	
Bottom Tertile	18879 (0.84)	3538 (0.16)
Mid Tertile	18455 (0.82)	3947 (0.18)
Top Tertile	18541 (0.83)	3842 (0.17)

Table B.14: Physical Activity by BMI Tertiles

BMI Tertiles	Physically Active	Not Active
	N (row %)	
Bottom Tertile	12251 (0.55)	10163 (0.45)
Mid Tertile	12023 (0.54)	10379 (0.46)
Top Tertile	10667 (0.48)	11735 (0.52)

Table B.15: Sedentary Lifestyle by BMI Tertiles

BMI Tertiles	Not Sedentary	Sedentary
	N (row %)	
Bottom Tertile	19378 (0.86)	3043 (0.14)
Mid Tertile	19321 (0.86)	3092 (0.14)
Top Tertile	18297 (0.82)	4114 (0.18)

B.2 Bivariate Tables for The Child Data

Table B.16: Mean Age by BMI Categories

BMI Categories	Age (in years)
	mean (SD)
Average	8.58 (4.40)
Overweight	9.29 (4.30)
Obese	9.56 (4.26)

Table B.17: Sex by BMI Categories

BMI Categories	Male	Female
	N (row %)	
Average	6026 (0.53)	5435 (0.47)
Overweight	2300 (0.50)	2346 (0.50)
Obese	1431 (0.50)	1410 (0.50)

Table B.18: Ethnicities by BMI Categories

BMI Categories	European Only	2+ Ethnicities (M)	Asian Only	Maori Only	Pacific Only	Other
	N (row %)					
Average	5041 (0.44)	2288 (0.20)	616 (0.05)	1481 (0.13)	1568 (0.14)	451 (0.04)
Overweight	1544 (0.33)	1075 (0.23)	186 (0.04)	891 (0.19)	527 (0.11)	422 (0.09)
Obese	612 (0.22)	626 (0.22)	79 (0.03)	705 (0.25)	273 (0.10)	545 (0.19)

Table B.19: Deprivation Quintiles by BMI Categories

BMI Categories	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
	N (row %)				
Average	1864 (0.16)	1975 (0.17)	2123 (0.19)	2506 (0.22)	2993 (0.26)
Overweight	537 (0.12)	619 (0.13)	777 (0.17)	1073 (0.23)	1640 (0.35)
Obese	175 (0.06)	245 (0.09)	366 (0.13)	630 (0.22)	1425 (0.50)

Table B.20: Urban/Rural Area by BMI Categories

BMI Categories	Urban	Rural
	N (row %)	
Average	4242 (0.83)	868 (0.17)
Overweight	1735 (0.84)	329 (0.16)
Obese	1077 (0.85)	184 (0.15)

Table B.21: Household Income by BMI Categories

BMI Categories	≤\$15,000	\$15,001–\$40,000	\$40,001–\$70,000	>\$70,000
	N (row %)			
Average	3977 (0.46)	367 (0.04)	1920 (0.22)	2307 (0.27)
Overweight	1355 (0.40)	186 (0.06)	891 (0.27)	915 (0.27)
Obese	621 (0.33)	131 (0.07)	580 (0.31)	535 (0.29)

Table B.22: Parents' Educational Qualifications by BMI Categories

BMI Categories	No Qualification	Secondary	Tertiary
	N (row %)		
Average	1709 (0.15)	4055 (0.37)	5280 (0.48)
Overweight	891 (0.20)	1782 (0.40)	1817 (0.40)
Obese	709 (0.26)	1205 (0.44)	827 (0.30)

Table B.23: Adherence to Fruit Guideline by BMI Categories

BMI Categories	No	Yes
	N (row %)	
Average	2620 (0.29)	6374 (0.71)
Overweight	1160 (0.31)	2533 (0.69)
Obese	789 (0.34)	1558 (0.66)

Table B.24: Adherence to Vegetable Guideline by BMI Categories

BMI Categories	No	Yes
	N (row %)	
Average	3885 (0.43)	5106 (0.57)
Overweight	1576 (0.43)	2114 (0.57)
Obese	1166 (0.50)	1184 (0.50)

Table B.25: Soft Drink Consumption by BMI Categories

BMI Categories	0/week	1/week	2–3/week	4+/week
	N (row %)			
Average	4005 (0.39)	2793 (0.27)	2196 (0.21)	1275 (0.12)
Overweight	1330 (0.32)	1096 (0.27)	1045 (0.25)	639 (0.16)
Obese	682 (0.28)	641 (0.26)	667 (0.27)	445 (0.18)

Table B.26: Fast Food Consumption by BMI Categories

BMI Categories	0/week	1/week	2–3/week	4+/week
	N (row %)			
Average	3161 (0.31)	5010 (0.49)	1856 (0.18)	257 (0.02)
Overweight	1148 (0.28)	1950 (0.47)	882 (0.21)	142 (0.03)
Obese	595 (0.24)	1103 (0.45)	632 (0.26)	113 (0.05)

B.3 Summary of Bivariate Statistical Tests

Table B.27: Bivariate Analysis of The Covariates and BMI Tertiles (Adult Data)

Covariates	<i>F</i>	df (between, within)	<i>p</i>	η^2	Effect Size
Age	355.25	2, 67728	<.001	.010	*

Covariates	χ^2	df	<i>p</i>	φ_c	Effect Size
Sex	875.59	2	<.001	.114	*
Ethnicity	6292.52	10	<.001	.216	**
Deprivation Quintile	2051.51	8	<.001	.123	*
Urban/Rural Area	72.99	2	<.001	.023	
Household Income	190.30	6	<.001	.038	
Education	969.21	4	<.001	.085	*
Fruit Guideline	47.09	2	<.001	.019	
Vegetable Guideline	26.27	2	<.001	.014	
Migration Status	848.29	2	<.001	.079	
Difficulty Climbing Several Flights of Stairs	1589.52	4	<.001	.108	*
Smoking Status	581.53	4	<.001	.066	
Drinking Problem	29.28	2	<.001	.015	
Physical Activity	261.00	2	<.001	.044	
Sedentary Lifestyle	253.08	2	<.001	.043	

F = F-statistic from ANOVA
 η^2 = Eta-squared
 * small effect size

χ^2 = Chi-squared statistic
 φ_c = Cramer's V
 ** medium effect size

Table B.28: Bivariate Analysis of The Covariates and BMI Categories (Child Data)

Covariates	<i>F</i>	df (between, within)	<i>p</i>	η^2	Effect Size
Age	82.44	2, 18945	<.001	.009	*

Covariates	χ^2	df	<i>p</i>	φ_c	Effect Size
Sex	14.05	2	<.001	.026	
Ethnicity	1390.14	10	<.001	.149	**
Deprivation Quintile	793.49	8	<.001	.112	*
Urban/Rural Area	4.61	2	.099	.011	
Household Income	150.86	6	<.001	.049	
Education	341.58	4	<.001	.074	*
Fruit Guideline	20.23	2	<.001	.022	
Vegetable Guideline	34.99	2	<.001	.029	
Soft Drink Consumption	183.59	6	<.001	.054	
Fast Food Consumption	134.29	6	<.001	.046	

F = F-statistic from ANOVA
 η^2 = Eta-squared
 * small effect size

χ^2 = Chi-squared statistic
 φ_c = Cramer's V
 ** medium effect size

C Proportional Odds Assumption

Table C.1: Odds Ratios from Binomial Regressions with Contrasts in The Adult Data

	$\frac{P(\text{Mid Tertile or Top Tertile})}{P(\text{Bottom Tertile})}$	$\frac{P(\text{Top Tertile})}{P(\text{Bottom Tertile or Mid Tertile})}$
	OR (95% CI)	
Deprivation Quintile (ref. group: Quintile 1)		
Quintile 2	1.08 (1.03–1.13)	0.94 (0.89–1.00)
Quintile 3	1.09 (1.04–1.15)	1.10 (1.05–1.16)
Quintile 4	1.25 (1.19–1.31)	1.33 (1.26–1.40)
Quintile 5	1.42 (1.35–1.50)	1.49 (1.41–1.58)
Linear Contrast	1.12 (1.07–1.17)	1.28 (1.22–1.33)
Ethnicity (ref. group: European only)		
2+ ethnicities (M)	1.71 (1.62–1.82)	1.67 (1.58–1.75)
Asian only	0.58 (0.54–0.62)	0.41 (0.38–0.45)
Māori only	2.96 (2.79–3.14)	2.84 (2.71–2.99)
Pacific only	9.50 (8.30–10.86)	7.50 (6.95–8.10)
Other	0.89 (0.85–0.93)	0.94 (0.89–0.99)
Age	1.02 (1.02–1.02)	1.003 (1.002–1.004)
Difficulty Climbing Several Flights of Stairs (ref. group: No difficulty)		
A little difficulty climbing stairs	1.61 (1.54–1.69)	2.01 (1.92–2.10)
A lot difficulty climbing stairs	1.56 (1.47–1.66)	2.02 (1.92–2.14)
Sex		
Female*	0.60 (0.58–0.61)	1.04 (1.01–1.08)
Smoking Status (ref. group: Non smoker)		
Ex smoker	1.41 (1.36–1.46)	1.34 (1.30–1.39)
Current smoker	0.92 (0.89–0.96)	0.92 (0.88–0.96)
Household Income (ref. group: ≤\$15,000)		
\$15,001–\$40,000	1.18 (1.11–1.25)	1.22 (1.15–1.30)
\$40,001–\$70,000	1.46 (1.37–1.56)	1.37 (1.28–1.46)
>\$70,000	1.69 (1.58–1.80)	1.49 (1.40–1.59)
Educational Qualification (ref. group: No qualification)		
Secondary	0.92 (0.88–0.96)	0.87 (0.83–0.91)
Tertiary	0.83 (0.79–0.87)	0.76 (0.73–0.79)
Physical Activity		
Not physically active	1.13 (1.09–1.16)	1.17 (1.14–1.21)
Migration Status		
Native	1.21 (1.14–1.28)	1.20 (1.12–1.29)

* Female had lower odds of being in the top two tertiles compared with being in the bottom tertile, but they had higher odds of being in the top tertile compared with being in the bottom two tertiles

Table C.2: Odds Ratios from Binomial Regressions with Contrasts in The Child Data

	$\frac{P(\text{Overweight or Obese})}{P(\text{Average})}$	$\frac{P(\text{Obese})}{P(\text{Average or Overweight})}$
	OR (95%CI)	
Deprivation Quintile (ref. group: Quintile 1)		
Quintile 2	1.12 (1.02–1.23)	1.27 (1.09–1.49)
Quintile 3	1.40 (1.28–1.53)	1.70 (1.47–1.96)
Quintile 4	1.49 (1.37–1.63)	1.82 (1.59–2.10)
Quintile 5	1.68 (1.54–1.82)	2.30 (2.01–2.64)
Linear Contrasts	1.15 (1.09–1.21)	1.19 (1.12–1.28)
Ethnicity (ref. group: European only)		
2+ ethnicities (M)	1.48 (1.39–1.57)	1.65 (1.51–1.80)
Asian only	1.03 (0.93–1.15)	1.12 (0.94–1.35)
Māori only	1.99 (1.86–2.14)	2.25 (2.05–2.48)
Pacific only	3.48 (3.18–3.81)	4.54 (4.08–5.06)
Other	1.19 (1.11–1.28)	1.33 (1.20–1.48)
Age	1.05 (1.04–1.05)	1.03 (1.02–1.04)
Parents' Qualification (ref. group: No qualification)		
Secondary qualification	0.74 (0.69–0.79)	0.74 (0.68–0.80)
Tertiary qualification	0.64 (0.60–0.69)	0.58 (0.53–0.63)
Soft Drink Consumption (ref. group: <1/week)		
1/week	1.15 (1.09–1.22)	1.22 (1.12–1.33)
2-3x/week	1.22 (1.15–1.30)	1.34 (1.23–1.46)
4+week	1.18 (1.10–1.26)	1.40 (1.27–1.54)
Sex		
Female	1.12 (1.07–1.17)	1.04 (0.98–1.11)

D Codebook

Table D.1: Codebook for Derived and Renamed Variables in This Thesis

Variable Name	Value Description
gender	Male, Female
bmscale	Body Mass Index in kg/m2
bmic	Average = a person with an average BMI, Overweight = an overweight person, Obese = an obese person
age	Age in years
euro	European ethnicity; 1 = Yes, 2 = No
maori	Māori ethnicity; 1 = Yes, 2 = No
pacific	Pacific ethnicity; 1 = Yes, 2 = No
asian	Indian and/or Chinese ethnicity; 1 = Yes, 2 = No
other	Ethnicity not included elsewhere; 1 = Yes, 2 = No
eth_count	Māori Only = a person who only belong to Māori ethnic group, Pacific Only = a person who only belong to Pacific ethnic group, European Only = a person who only belong to European ethnic group, Asian = a person who belong to Asian and/or Indian ethnic group, 2+ Ethnicities (M) = a person who belong to multiple ethnicities with Māori ancestry, Other = a person who belong to multiple ethnicities or other ethnicities not elsewhere included.
native	Migrant = a person who moved to settle in NZ less than 10-11 years ago, Native = New Zealanders and those who moved to settle in NZ more than 10-11 years ago
secondary	1 = no qualification, 2 = secondary qualification
tertiary	0 = no tertiary qualification, 3 = tertiary qualification
edu	No Qualification = no educational qualification, Secondary = secondary educational qualification (i.e. National Certificate level 1 to 2, NZ University Entrance before 1986 in one or more subjects, NZ Higher School Certificate, University Entrance qualification, NZ A or B Bursary; Scholarship; or National Certificate Level 3, Other NZ secondary school qualification, and Overseas secondary school qualification), Tertiary = tertiary educational qualification (i.e. Bachelor's degree, Bachelors degree with honours, MA, Msc, PhD, Diploma, Diploma - Post Graduate, Trade or technical certificate which took more than 3 months full time study, Professional qualifications, and other tertiary qualifications)
hhinc	Household income per year: <=\$15,000; \$15,001–\$40,001–\$70,000; >\$70,000
dep	NZ deprivation quintile; 1 = quintile 1, 2 = quintile 2, 3 = quintile 3, 4 = quintile 4, 5 = quintile 5; 2002/03 NZHS used deprivation quintile data from the 2002 Census, 2006/07–2012/13 NZHS used deprivation quintile data from the 2006 Census, 2013/14–2014/15 NZHS used deprivation quintile data from the 2013 Census
UA	Urban = an area with at least a 1000 population, Rural = an area with less than 1000 population
haz_drinker_all	Alcohol Problem = AUDIT score >=8, No Alcohol Problem = AUDIT score <8
minutes	duration of physical activity in the past week in minutes (rigorous physical activity is multiplied by two)
active	Active = a person who spent 30 minutes of brisk walk or moderate physical exercise (15 minutes if it was a vigorous physical activity) in the past week, Not Active = a person who spent less than 30 minutes of brisk walk or moderate physical activity (less than 15 minutes if it was vigorous physical activity) in the past week
sedentary	Sedentary = a person who had not had at least 30 minutes of exercise in the past week, Not Sedentary = a person who had at least 30 minutes of exercise in the past week.
stair	A Lot Difficult = a person who had a lot difficulties climbing several flights of stairs, A Little Difficult = a person who had a little difficulties climbing several flights of stairs, No Difficulty = a person who had no difficulty climbing several flights of stairs
fruit	Yes = a person who met 2+ servings of fruit per day, No = a person who had less than 2 servings of fruit per day
veges	Yes = a person who met 3+ servings of vegetable per day, No = a person who had less than 3 servings of vegetable per day
softd	Soft drink consumption: 0/week, 1/week, 2–3/week or 4+/week
fastf	Fast food consumption: 0/week, 1/week, 2–3/week or 4+/week
cluster	Primary Sampling Unit (PSU) Indicator
strata	Stratum Indicator
finalwgt	Final sample weight